Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Improved Stereo Matching Technique for Digital Elevation Model Generation in Mountainous Terrain

Nidhi Dubey

dubey.nidhi52@gmail.com

Faculty of Engineering Sciences and Technology, Adani University

Hitesh Chhinkaniwala

Faculty of Engineering Sciences and Technology, Adani University

Article

Keywords: Digital Elevation Model, Stereo Satellite Imagery, Rational Polynomial Coefficients, Ground Control Points, Deep Learning, Terrain Analysis

Posted Date: March 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8946063/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Improved Stereo Matching Technique for Digital Elevation Model Generation in Mountainous Terrain

Nidhi Dubey[1*] and Hitesh Chhinkaniwala[1]

[1]Faculty of Engineering Sciences and Technology, Adani University, Ahmedabad, India.

*Corresponding author(s). E-mail(s): dubey.nidhi52@gmail.com;
Contributing authors: hitesh.chhinkaniwala@adaniuni.ac.in;

Abstract

Geospatial applications like terrain analysis, infrastructure design, and area mapping require accurate Digital Elevation Models (DEMs) derived from satellite stereo imagery. Classical photogrammetric approaches using Rational Polynomial Coefficients (RPCs) and Ground Control Points (GCPs) remain widely adopted for precise elevation modeling. Recently, Deep Learning (DL) based stereo matching methods have emerged as promising alternatives, enabling automated DEM extraction in complex terrain conditions. The proposed work presents a systematic, application-oriented comparison between conventional photogrammetric workflows and DL assisted stereo pipelines for DEM generation using CARTOSAT-1. Quantitative assessment was conducted against high-quality reference data using standard accuracy metrics, including Root Mean Square Error (RMSE), Mean Absolute Error (MAE), terrain-specific error statistics, cross-sectional elevation profile analysis, and terrain error mapping. Results indicates, incorporating GCPs significantly improves geometric accuracy, achieving vertical RMSE values within a few meters. DL based stereo approaches that operates w/o GCPs, demonstrate competitive performance in overall elevation estimation. Among the evaluated networks, GANet achieves the lowest error and highest structural consistency, especially in steep and vegetated regions. While RPC-based photogrammetric methods with GCPs provide the highest absolute accuracy. DL shows great potential for automated DEM generation where GCPs are scarce, highlighting the complementary strengths of both approaches.

Keywords: Digital Elevation Model, Stereo Satellite Imagery, Rational Polynomial Coefficients, Ground Control Points, Deep Learning, Terrain Analysis

1 Introduction

Digital Elevation Models (DEMs) accurately represent the earth's surface, enabling geospatial applications like urban planning, hydrological modeling, disaster risk assessment, infrastructure development, and environmental monitoring. High-resolution satellite images are now readily accessible at lower costs, improving scale-effectiveness. Stereo photogrammetry is proven as a viable solution to large-scale production of DEM. CARTOSAT-1's along-track stereo imaging capability and good geometric configuration have made it widely used in regional terrain mapping[1]. Traditionally, satellite stereo-based DEM generation is based on strict photogrammetric principles; sensor geometry modelled via RPCs and can be further optimised with GCPs to achieve greater levels of geometric accuracy[2, 3]. Techniques of image matching, e.g. Semi Global Matching (SGM) have been extensively used because of the better tradeoff between performance and accuracy [4]. These methods have good results in well-textured and averagely undulating landscapes, but they fail in adverse conditions, such as low-texture areas, occlusions, shadows, and steep landscapes, resulting in mismatches and elevation artifacts[5].

Stereo matching has largely evolved due to recent developments in the field of DL techniques, which makes it possible to estimate the disparity between two images based on information in local texture areas and in global contexts. Convolutional Neural Network (CNN) based models, such as Global Context Network (GCNet) and Pyramid Stereo Matching Network (PSMNet), are proven to be more effective than conventional models in generating dense and smooth disparity maps, especially in complex terrains [6, 7]. Recent hybrid systems combining transformers, recurrence-refined modules, and semantic constraints have improved disparity estimation accuracy for high-resolution satellite images [8, 9]. The proposed learning-based methods provide strong alternatives to manually constructed similarity measures, particularly in a heterogeneous topography where conventional photogrammetric techniques are likely to be ineffective.

Regardless of these developments, the implementations of DL-based stereo matching into the working patterns of operational satellite photogrammetry are still an area of research. Numerous studies done so far focus predominantly on the accuracy of disparities and the quality of visual representation. It is observed that less effort has been devoted to the assessment of DL-based DEMs with varied orientation schemes, i.e., RPCs-w/o-GCPs and RPCs-with-GCPs, which is important to determine their suitability to real-world mapping systems [10].

This paper compares DEM generation using CARTOSAT-1 stereo imagery with traditional photogrammetry and DL-based stereo matching. DEM accuracy (with or w/o GCPs from RPCs) was assessed against reference DEMs via quantitative error analysis, terrain-based assessment, profiling, and spatial error mapping. Key contributions include:

1. A DEM generation pipeline equipped with CARTOSAT-1 stereo photogrammetry and DL, used to systematically evaluate the performance of DEM using various geometries and GCP refinement, and using performance dependent on terrain characteristics and evaluates traditional versus DL.

2. An integrated platform: to compare the traditional architecture and DL architecture on various terrains and images settings.
3. Benchmarking of classical photogrammetric and DL-based stereo pipelines under identical circumstances, with respect to application-oriented comparison instead of the new theory.

2 Literature Survey

Generating DEMs from stereo imagery has long been crucial in photogrammetry and remote sensing. The growing availability of high-resolution satellite and UAV imagery has made automated and accurate DEM production essential.

2.1 DEM Generation from Stereo Imagery

Traditional DEM generation systems rely on rigorous geometric modelling and handcrafted similarity measures to establish correspondence between stereo image pairs. Classical pipelines typically include image orientation, epipolar rectification, dense stereo matching, and triangulation, often based on camera geometry models such as RPCs. Algorithms such as SGM, Census Transform, and Normalized Cross-Correlation (NCC) have been widely adopted due to their robustness and geometric consistency [4, 11]. These methods are advantageous because they have a strong theoretical foundation and do not require training data, making them reliable baseline approaches for DEM generation. However, their performance degrades in low-texture regions, repetitive patterns, shadowed areas, and complex mountainous terrains, which are common in high-resolution satellite imagery [8]. Traditional photogrammetric approaches emphasize explicit geometric constraints and deterministic matching strategies. While these methods ensure high geometric fidelity, their reliance on handcrafted similarity metrics limits adaptability under challenging imaging conditions. As a result, their robustness and completeness are often constrained in heterogeneous landscapes, as discussed in Table 1.

Table 1: Traditional Stereo Matching Methods for DEM Generation

Method	**Core Principle**	**Advantages**	**Limitations**
NCC	Intensity correlation	Simple and interpretable	Sensitive to radiometric variations
Census	Non-parametric transform	Robust to illumination changes	Poor performance in homogeneous regions
SGM	Pathwise cost aggregation	Good accuracy–speed trade-off	Fails in occlusions and high-relief terrain
RPCs-with-GCPs	Rigorous sensor geometry	High geometric accuracy	Requires manual GCPs collection

2.2 DL–Based Stereo Matching

The advent of DL introduced a paradigm shift in stereo matching by replacing handcrafted similarity measures with learnt feature representations. Early CNN-based architectures demonstrated that neural networks could effectively learn correspondence patterns and outperform classical methods, particularly under challenging conditions. Models such as GCNet, PSMNet, and GWCNet introduced cost-volume construction and 3D convolutional regularisation, enabling improved disparity estimation through global context modelling [6, 7]. These architectures achieved strong performance in both textured and non-textured regions and laid the foundation for learning-based DEM generation pipelines. Earlier most deep stereo networks have been trained on automotive datasets, but now these networks have been applied to satellite and aerial imagery. All these adaptations are needed because remote sensing data have such peculiarities as high image sizes, extreme parallax, variability in radiometrics, and change in domain. Empirical evidence demonstrates the superiority of DL models over SGM when it comes to the completeness and structural consistency of DEM, especially in the urban and mountainous terrains [12, 13]. The semantic-aware stereo networks can also be combined with other disparity estimations, including SemStereo, which offers a difference in SemStereo with consideration of the scene-level data in difficult regions. A major challenge in applying DL stereo models to DEM generation is the dataset shift. Models trained on synthetic or near-range imagery often exhibit reduced performance when applied to satellite data. Cross-domain evaluations over high-relief terrain resulted in performance degradation exceeding 40% [13]. To address this issue, recent research has focused on domain adaptation, self-supervised learning, and terrain-aware sampling strategies [14]. These approaches highlight the importance of dataset-specific training protocols and standardised evaluation frameworks for reliable DEM generation, depicted in Table 2.

Table 2: DL–Based Stereo Networks for DEM Generation

Model	Proposed Approach	Application Domain	Major Limitation
GCNet	3D cost volume construction	General stereo vision	High memory cost
PSMNet	Spatial pyramid pooling	Satellite and aerial imagery	Domain sensitivity
GANet	Guided aggregation strategy	Urban and vegetated terrain	Computationally heavy
SemStereo	Semantic consistency constraints	Remote sensing imagery	Requires semantic labels
Transformer-based models	Global self-attention mechanism	High-resolution satellite imagery	Large training data demand

2.3 Integrating Photogrammetry and DL

Recent research has increasingly focused on hybrid approaches that integrate traditional photogrammetric geometry with DL–based disparity estimation. In such pipelines, RPCs-based orientation ensures geometric accuracy, while DL models provide dense and structurally consistent disparity maps [9]. This integration leverages the complementary strengths of both paradigms—geometric rigor from classical photogrammetry and adaptability from data-driven models—resulting in improved robustness and operational reliability for DEM generation [13]. There have been serious advances in the field of DL-based stereo matching, but it usually considers single architecture or no universal protocols on DEM generation. Further, the extant literature lacks systematic comparisons of DL models within consistent pre-processing and triangulation photogrammetric parameters as shown in Table 3.

Table 3: Comparison of Traditional vs. DL Methods for DEM Generation

Aspect	**Traditional Methods**	**DL Methods**
Geometry	Explicit and rigorous	Implicitly learned
Training data	Not required	Required
Performance in low-texture regions	Limited	Strong
Level of automation	Moderate	High
Generalization capability	Strong	Dataset dependent
DEM completeness	Moderate	High

3 Dataset Description

3.1 Study Area and CARTOSAT-1 Data

The Space Application Centre (SAC), Indian Space Research Organisation (ISRO) uses CARTOSAT-1 along-track stereo to assess and compare typical photogrammetric and DL-based generation of DEM [1, 3]. The sensor comprises two panchromatic cameras (FORE and AFT) with view angles of 26° at nadir and allow stereo imaging to be made within almost the same illumination levels [11].

The sources, where the selected datasets are found, are heterogeneous terrain conditions, such as urban territories, vegetated areas, and agricultural plains as well as mountainous areas with steep slopes. This variation is necessary to measure stereo matching performance between the validity of different surface textures and surface elevation gradients [5, 8]. The spatial resolution of the panchromatic image is around 2.5 m; thus, it can be used for DEM generation at a regional level and commonly used in mapping. The CARTOSAT-1 stereo pairs come with vendor-supplied RPCs, detailing the geometry of the sensors, facilitating photogrammetric orientation, epipolar rectification, and 3D reconstruction [3, 11], shown in Table 4. Other metadata, such as acquisition parameters and sensor calibration data, were also used so that

proper preprocessing would be done. Qualified reference DEMs were used in the accuracy assessment, which were represented by LiDAR surveys or proven high-resolution DEM products [5, 10]. To reduce systematic spatial offsets, these reference datasets were co-registered with the CARTOSAT-1 imagery and had high vertical accuracy.

3.2 Pre-processing and Stereo Rectification

Before the generation of DEM, conventional preprocessing was provided on the stereo imagery. The radiometric normalisation was carried out to diminish the difference in brightness between the FORE and the AFT images. Epipolar rectification was then applied with respect to the RPCs-based orientation models which brought the stereo pairs into epipolar line correspondence with each other [3, 11]. Both the traditional dense image matching algorithms, including SGM [4] and DL stereo networks, trained to directly estimate disparity, processed the rectified stereo images. [6, 7].

3.3 Orientation Strategies and Ground Control Points

There were two orientation strategies followed to examine the effect on geometric refinement on DEM of accuracy. The former used only RPC-based orientation, and the latter included few GCPs that were accurately surveyed to counter systematic RPCs biases [3, 10]. Evenly spaced, the GCPs had a diverse elevation of the study area to enhance geometric stability. This architecture allows us to independently assess the disparity estimation of DL under various photogrammetric accuracy conditions, which is consistent with the findings of CARTOSAT-1 in the past [5, 8].

3.4 DEM Reconstruction and Evaluation

In the context of DL, fixed-size patches of the rectified stereo images were used to train them on the high performance computing system [6, 9]. Disparity space generated them through forward intersection of the reference DEMs projected into disparity space in RPCs [8]. The dataset was divided into training, validation and testing parts at the scene level to prevent space overlap. The training samples comprised a variety of land cover and terrain to further improve model generalization to different conditions of surface and elevations [12].

The disparity maps of both conventional algorithms and DL models have been transformed into elevation values through RPCs-based triangulation, and this transformed such data were used to produce CARTOSAT-1-like DEMs [8, 11]. A similar spatial resolution of 10 m was used to resample all DEMs so that they are consistent in evaluation [5]. The assessment framework encompasses the quantitative comparison with reference DEMs and terrain-based accuracy test during steep hillsides, vegetation, and built-ups, cross profile analysis of the texture to determine structure consistency; and the spatial error layer to pinpoint local high or low displacement of the elevation. This standardization of the evaluation strategy allows comparing the traditional and DL-based DEM generation systems rigorously, as shown in Table 3.

Table 4: Summary of CARTOSAT-1 Datasets Used in This Study

Dataset Type	**Source**	**Spatial Resolution**	**Role in Workflow**
CARTOSAT-1 FORE & AFT stereo images	SAC, ISRO	∼2.5 m	Stereo matching and disparity estimation
RPCs metadata	SAC, ISRO	–	Sensor geometry and triangulation
GCPs	Survey / national datasets	–	RPCs bias compensation
Reference DEM	LiDAR / validated DEM	1–5 m	Ground truth for accuracy evaluation
Terrain masks	Derived from reference DEM	10 m	Terrain-wise error analysis

4 Methodology

The proposed work adopts a two-stage evaluation framework that integrates conventional photogrammetric processing with DL–based stereo matching for DEM generation. Such hybrid strategies are increasingly adopted to combine the geometric rigour of photogrammetry with the representational power of deep stereo networks [8, 9, 15], depicted in Fig 1.

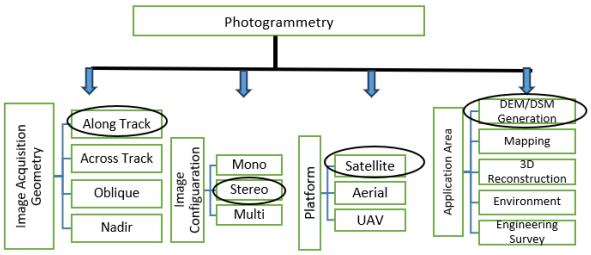


Fig. 1: Classification of photogrammetry based on image acquisition geometry, image configuration, sensing platform, and application domains.

The proposed framework evaluates geometric accuracy, disparity-level precision, and terrain sensitivity across methods. Performance is assessed using RMSE, MAE, and geometric fidelity metrics, consistent with established practices in satellite stereo analysis [15–17].

4.1 DEM Generation and Evaluation Pipeline

RPCs are commonly used to model the sensor geometry of satellite images and undergo preprocessing and orientation before being utilized for DEM reconstruction. RPC-based orientation is a well-established technique in satellite photogrammetry [5, 15]. To further improve orientation accuracy, GCPs are incorporated into the RPC refinement process [5, 16].

Traditionally, DEM generation is performed through epipolar rectification followed by dense stereo matching, which serves as the reference workflow in satellite photogrammetry [5, 8]. In parallel, recent deep learning–based approaches employ networks such as UNet, PSMNet, and GANet to directly estimate dense disparity maps, which are subsequently converted into DEMs. State-of-the-art stereo networks have demonstrated reduced disparity errors on satellite imagery, particularly in regions with complex terrain [8, 9, 15, 18].

4.2 RPCs-Based Photogrammetric Orientation (w/o GCPs)

4.2.1 RPCs Sensor Geometry

RPCs provide a generalised sensor model that relates ground coordinates (X, Y, Z) to image coordinates (r, c) using normalised polynomial functions [3, 19, 20]. The RPCs model is expressed as

$$(r, c) = \left(\frac{Q_1(X, Y, Z)}{P_1(X, Y, Z)}, \qquad \frac{Q_2(X, Y, Z)}{P_2(X, Y, Z)}\right) \tag{1}$$

where P_1, P_2, Q_1, and Q_2 are third-degree polynomials comprising up to 78 coefficients. Normalisation using offset and scale parameters ensures numerical stability.

The modern satellite missions include WorldView, GeoEye, Pleiades and ZY-3 which include RPCs metadata with imagery, making it possible to orientate directly towards photogrammetry even without physical sensor models [8, 15]. RPCs-w/o-GCPs orientation does not use any external control to generate the DEM, but is vulnerable to systematic distortions and terrain-specific distortions, such as those found in rough terrain [3, 5].

4.2.2 RPCs-w/o-GCPs DEM Generation Workflow

RPCs parameters are normalised and used to establish forward and inverse projections between image and ground space [5, 15]. Stereo rectification is performed in the epipolar domain, aligning conjugate points along horizontal scanlines [9].

Disparities are estimated using similarity measures, and elevation is computed through RPCs-based triangulation:

$$Z = f(d, \text{RPCs}) \tag{2}$$

where d denotes pixel disparity derived from stereo correspondences [8].

4.2.3 Limitations of RPCs-w/o-GCPs Orientation

RPCs-w/o-GCPs orientation is computationally inexpensive and it has vertical bias and geometric drift because it has no point of reference [5]. Large-scale and steep or heterogeneous terrain as well as similar terrain, exaggerate errors which may require refinement or post-processing to produce an accurate DEM [8, 15].

4.3 Refinement of RPC-Based Sensor Orientation Using GCPs

4.3.1 RPCs Refinement Using GCPs

GCPs provide geodetically accurate references to compensate systematic biases in vendor-supplied RPCs [5, 8]. Corrected image coordinates (r', c') are obtained using bias correction functions estimated from GCPs residuals:

$$(r', c') = (r + \Delta r(X, Y, Z), \quad c + \Delta c(X, Y, Z)) \tag{3}$$

4.3.2 GCPs-Based Refinement Workflow

GCPs are acquired from high-accuracy reference datasets, and residuals between observed and RPCs-projected image coordinates are computed. These residuals are modelled using low-order correction functions to refine the RPCs [3]. Stereo rectification and triangulation using refined RPCs yield DEMs with improved geometric consistency and terrain representation [8, 9].

4.3.3 Benefits of RPCs Optimisation

RPCs refinement minimises systematic distortion and the accuracy of elevation significantly, especially over complicated topography [5, 10]. It is related to the improvement of the geometric consistency of the gradients as well as the more life-like depiction of steep slopes, making the optimisation of the RPCs a vital preprocessing instrument of the stable production of DEM [9, 15].

4.4 DL–Based DEM Generation

DL stereo methods estimate disparity directly from image data by learning hierarchical features and global contextual cues, eliminating handcrafted similarity measures [8, 9, 18]. All models follow a unified pipeline comprising stereo image input, dense disparity prediction, and RPCs-based triangulation for DEM reconstruction as referred to above equation. Recent studies consistently report superior performance of DL stereo methods over conventional approaches, particularly in low-texture, shadowed, and vegetated regions [13, 21]. To ensure reproducibility and fairness, all models are trained and evaluated using identical data splits, augmentation strategies, optimisation parameters, and computational environments [12].

The deep learning–based disparity estimation framework employed in this study integrates three representative stereo architectures- UNet, GANet, and PSMNet, selected to capture complementary strengths in dense matching for high-resolution remote sensing imagery. The UNet-based model follows a classical encoder–decoder

Table 5: Training Configuration and Computational Setup for DL Stereo Models Applied to CARTOSAT-1 Along-Track Stereo Imagery in Hilly Terrain (Uttarakhand, India)

Parameter	**UNet**	**PSMNet**	**GANet**
Stereo Dataset	CARTOSAT-1 Along-Track Stereo Images (Uttarakhand, India)		
Terrain Characteristics	High relief, steep slopes, shadowed and vegetated regions		
Input Patch Size	256×256	256×256	256×256
Maximum Disparity	192 px	192 px	192 px
Trainable Parameters	$\sim$11.8 M	$\sim$5.5 M	$\sim$34.6 M
Cost Volume Construction	Implicit	Explicit 4D	Explicit 3D
Aggregation Strategy	Encoder–decoder	Stacked hourglass	Guided aggregation
Loss Function	Smooth L1	Smooth L1	Smooth L1 + Soft ArgMin
Optimizer	Adam	Adam	Adam
Initial Learning Rate	1×10^{-4}	1×10^{-3}	1×10^{-4}
Learning Rate Schedule	Step decay	Cosine annealing	Fixed
Batch Size	8	4	2
Training Epochs	60–80	100	80–100
GPU Used	RTX 3090 (24 GB)	RTX 3090 (24 GB)	RTX 3090 (24 GB)
Training Time / Epoch	$\sim$3.5 min	$\sim$6 min	$\sim$11 min
Total Training Time	$\sim$5–6 h	$\sim$10 h	$\sim$12–15 h
DEM Reconstruction	RPC-based	RPC-based	RPC-based

design with skip connections, enabling effective preservation of fine-scale texture and hierarchical spatial features, which are critical for accurate elevation reconstruction. GANet, a guided-learning stereo network, constructs a three-dimensional cost volume and applies guided semi-global and local aggregation with 3D sparse convolutions and attention mechanisms, yielding robust performance under challenging conditions such as shadows, occlusions, and low-texture or vegetation-covered regions. In contrast, PSMNet leverages spatial pyramid pooling and stacked hourglass modules to capture multi-scale contextual information through a four-dimensional cost volume, facilitating iterative refinement of disparity estimates by integrating local and global correspondence cues. All three models are trained using Smooth L1–based loss functions and Adam optimization with architecture-specific learning rate schedules and batch sizes. The detailed training configurations, computational setup, and dataset-specific parameters for CARTOSAT-1 stereo imagery over hilly terrain are summarized in Table 5.

5 Results and Analysis

5.1 Evaluation of DEMs Using RPCs-w/o-GCPs Orientation

DEMs were produced using traditional photogrammetric methods (RPCs-based orientation with and w/o GCPs) and modern DL-based stereo matching (UNet, PSMNet,

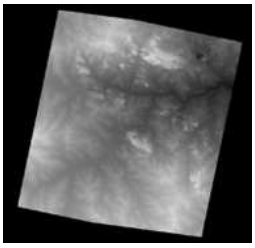
and GANet) techniques and compared. These assessment criteria are positional accuracy, convergence properties, terrain performance, disparity-level fidelity, and model fidelity. The findings encompass the traditional orientation-based DEM generation and the modern DL models to deliver the same insights into DEM behaviour in heterogeneous terrains.

DEMs obtained on the Hyderabad region with the RPCs-based orientation are demonstrated in Fig 2a. These DEMs resulted out of CARTOSAT-1 stereo imagery (with a resolution of 2.5 m) is a representation of a geologically complex place that is typified by rough granite terrains and solitary hillocks. There are high errors in the RPCs-w/o-GCPs DEMs in and around terrain discontinuities, steep slopes, and urban structures. These errors are caused mainly by the absence of external geometrical constraints, which cause elevation bias, horizontal drift and local irregularities. RPCs-w/o-GCPs DEMs have quantitatively high error levels: RMSE of 50–70 m, MAE of about 60 m, and Standard Deviation of 10–15 m.

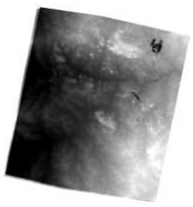
The goal is to determine the accuracy of DEM when it is used along with RPCs as well as when it is applied with RPCs using GCPs. The relative position and elevation difference were calculated against the CARTOSAT-1/SRTM reference DEM. The RPM modelled on satellite-based RPCs was used as the baseline orientation. Further enhancement was done by adding GCPs. DEMs were all tested at a resolution of 10 m to examine geometric fidelity and maintenance of the morphology of a terrain.

5.2 DEM Evaluation Using RPCs-with-GCPs Orientation

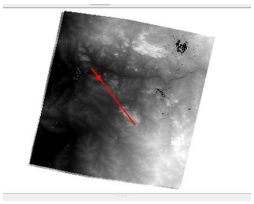
DEMs developed by using RPCs-w/o-GCPs and those developed by using RPCs-with-GCPs are compared in Fig 2b. Combination of GCPs leads to a significant increase in alignment, minuscule vertical offsets, and enhanced depiction of slopes and geomorphic buildings. As shown in Fig 2c the GCP-integrated DEMs are performing better, especially on steep landforms and when there is vegetation or surface discontinuities. The photogrammetric solution is tied to geodetically accurate ground points with the addition of GCPs, and therefore, drift is minimised, elevation bias is reduced, and vertical consistency is enhanced.



(a) DEM generated using RPCs (without GCPs)



(b) DEM generated using RPCs with GCPs



(c) Overlapping DEMs (RPCs vs RPCs-with-GCPs)

Fig. 2: Comparison of DEM generation approaches and their elevation agreement.

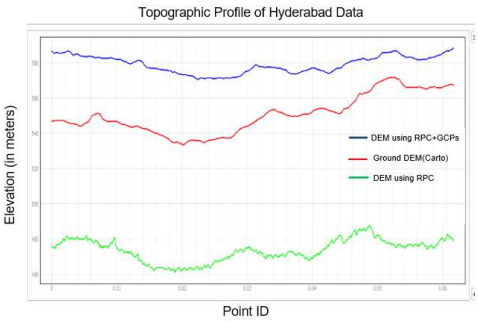


Fig. 3: Cross-sectional profile comparison on Hyderabad region data.

Table 6: Accuracy comparison of DEM generation methods using RPCs-based orientation.

DEM Method	**RMSE (m)**	**MAE (m)**	**Std. Dev. (m)**
RPCs-with-GCPs	2–3	∼2	<3
RPCs-w/o-GCPs	50–70	∼60	10–15

The comparison of topographic profiles between RPCs-w/o-GCPs DEM, RPCs-with-GCPs DEM and the reference CARTOSAT-1 DEM (Fig 3 confirms that, RPCs-with-GCPs DEM is closest to the ground truth with a significant improvement of up to 60-70 m, as shown in Table 6.

5.3 DL–Based DEM Generation Results

DL models analyze stereo satellite data of Uttarakhand, a mountainous region with dense vegetation and frequent cloud cover, for the second stage of testing. All models were trained to generate disparity maps, which were subsequently converted into DEMs. The analysis includes convergence behavior, disparity-level accuracy, structural similarity, robustness across terrain types, and DEM accuracy.

5.4 DL Model Performance Summary

The photogrammetric analysis proved in above section is that RPCs-w/o-GCPs DEMs give a very high amount of elevation errors, whereas RPCs-with-GCPs DEMs give highly accurate results (23 m). DL models attain DEM accuracies between 3.5 m–4.7 m w/o GCPs and as such, they have great potential in automated extraction of DEM. GANet is one of the DL methods that created more smoothly structured and coherent DEMs in high-relief variation regions, but remained slightly less accurate compared to the RPCs-with-GCPs photogrammetric answers. One potential avenue of research is that DL-based disparity estimation should be combined with photogrammetric constraints like RPCs-based block adjustment that integrates the merits of the two methods.

The disparity-level and elevation-level accuracy trends across the evaluated models are illustrated in Table 7. The consistent reduction in both disparity RMSE and DEM RMSE from UNet to GANet indicates that improvements at the stereo matching stage directly translate into enhanced elevation accuracy. GANet achieves the lowest disparity RMSE (2.51 px), which propagates into the lowest DEM RMSE (3.5 m), confirming the effectiveness of guided cost aggregation in preserving depth continuity over complex terrain as discussed in Table 8. In contrast, higher disparity errors

Table 7: Consolidated performance comparison across models and terrain types.

Category	Metric	UNet	PSMNet	GANet	Remark
Dataset	Region	Uttarakhand			Himalayan terrain
	Imagery	CARTOSAT-1			Along-track stereo
Architecture	Type	Enc–Dec	3D CNN	Guided Agg.	GANet robust
	Strength	Fast	Global	Edge-aware	Best in complex terrain
	Limitation	Slope sens.	Long train	High comp.	All terrain-sensitive
Training	Epochs	65	100	85	UNet fastest
Disparity	MAE (px)	1.82	1.35	1.21	GANet best
	RMSE (px)	3.45	2.78	2.51	Most consistent
	SSIM	0.91	0.94	0.95	Best structure
DEM	RMSE (m)	4.7	3.9	3.5	Highest fidelity
Terrain	Slopes	High	Med.	Low	GANet superior
	Vegetation	High	Mod.	Best	Foliage robust
	Shadows	High	Med.	Low	Better occlusion
Overall	Rank	III	II	I	GANet best

observed in the UNet lead to amplified elevation inaccuracies, particularly in regions with steep slopes and surface discontinuities.

Table 8: Accuracy metrics, terrain influence, and overall performance ranking.

Category	Metric	UNet	PSMNet	GANet
Disparity accuracy	MAE (px)	1.82	1.35	1.21
	RMSE (px)	3.45	2.78	2.51
	SSIM	0.91	0.94	0.95
DEM accuracy	DEM RMSE (m)	4.7	3.9	3.5
Terrain influence	Steep slopes	High	Medium	Low
	Vegetation	High	Moderate	Best
	Shadows	High	Medium	Low
Overall ranking	Rank	III	II	I

6 Conclusion

The paper proposed an end-to-end, geometry-consistent framework for generating DEMs from CARTOSAT-1 along-track stereo images. The proposed work combining conventional photogrammetric processing with the latest DL based methods. Majority of the work done so far is to separate and evaluate the accuracy of the stereo

correspondence or disparity; the proposed work evaluates the entire chain of DEM generation, RPCs orientation, GCPs refinement, stereo matching, DEM reconstruction and terrain-stratified accuracy evaluation. This ascertains the level of improvements within the discrepancy level into a significant level of accuracy on elevations. It has been established through the experimental study that vendor-supplied RPCs cannot be relied upon to generate DEM in high-relief terrain, which causes significant elevation bias and spatial residual. This is because the incorporation of a few GCPs substantially increases geometric fidelity, which reduces DEM errors by more than an order of magnitude and even eliminates elevation profile consistency with reference data. These results support the further applicability of the strict photogrammetric orientation as an accuracy measure of DEM, especially in mountainous conditions.

The stereo matching methods trained by DL possess high potential for providing automated generation of DEM despite the lack of GCPs. GANet always performs better on the measures of RMSE, MAE, and structural similarity, particularly at steep, vegetated, and shadowed areas where classical methods tend to fail. The terrain-wise analysis and spatial error mapping support better surface continuity and a smaller number of local artifacts, which signify the strength of deep stereo networks in unfavourable imaging conditions. The DEMs based on learning are not currently as vertical as the RPCs-with-GCPs of photogrammetric solutions. The result emphasises that deeper learning is not meant to be considered an alternative to photogrammetric modelling but a complementary element that could for example, be helped by clear geometric constraints.

Domain adaptation and self-supervised methods must be further extended to enhance proper generalisation of the study to a wide variety of terrain and imaging scenarios. The possibility of extending the suggested framework to multi-sensor fusion, which includes LiDAR or SAR data, is a promising direction for further improving the accuracy of DEM and the robustness of its operation.

7 Funding

The proposed research work received no external funding.

8 Conflicts of interest

The authors claim that they have no financial interests, commercial affiliations, or other possible conflicts of interests that may have impacted the objectivity of the given research nor have written this paper.

9 Data Availability Statement

The CARTOSAT-1 imagery analysed in this study was obtained under the license from the National Remote Sensing Centre (NRSC), ISRO, and is not publicly available. The data is available on the Bhoonidhi portal of ISRO with institutional authorization and in accordance with the data access policies—https://bhoonidhi.nrsc.gov.in. Interested researchers can apply for data through the official Bhoonidhi request mechanism, following ISRO's data dissemination guidelines.

References

[1] Srivastava, P.K., Saxena, R., Mukherjee, S.: Cartosat-1: A technical overview. Current Science **93**(5), 622–627 (2007)

[2] Toutin, T.: Review of geometric correction of remote sensing images using rational functions. ISPRS Journal of Photogrammetry and Remote Sensing **59**, 111–130 (2004)

[3] Fraser, C.S., Hanley, H.: Bias-compensated rpcs for sensor orientation of high-resolution satellite imagery. Photogrammetric Record **20**(109), 49–63 (2005)

[4] Hirschmüller, H.: Stereo processing by semi-global matching and mutual information. IEEE Transactions on Pattern Analysis and Machine Intelligence **30**(2), 328–341 (2008)

[5] Gong, K., Fritsch, D.: A detailed study about digital surface model generation using high resolution satellite stereo imagery. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences **III-1**, 69–76 (2016)

[6] Chang, J.-R., Chen, Y.-S.: Pyramid stereo matching network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5410–5418 (2018)

[7] Kendall, A., Martirosyan, H.: End-to-end learning of geometry and context for deep stereo regression. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 66–75 (2017)

[8] He, X., Jiang, S., He, S., Li, Q., Jiang, W., Wang, L.: Deep learning-based stereo matching for high-resolution satellite images: A comparative evaluation. ISPRS Archives **XLVIII-1/W2**, 1635–1642 (2023)

[9] Yang, M., Jiang, S., Jiang, W., Li, Q.: Stereo matching network with transformer-cnn feature fusion for satellite images. ISPRS Annals **X-G**, 995–1002 (2025)

[10] Pelmo, Y., Dorji, H., Dorji, P.: Accuracy assessment of dem generated from geoeye-1 stereo imagery using gcp-based rpc refinement. Journal of the National Engineering College **10**, 23–33 (2021)

[11] Toutin, T.: Review of geometric correction of remote sensing images using rational functions. ISPRS Journal of Photogrammetry and Remote Sensing **59**, 111–130 (2004)

[12] Wu, T., Vallet, B., Pierrot-Deseilligny, M., Rupnik, E.: An evaluation of deep learning based stereo dense matching dataset shift from aerial images and a large scale stereo dataset. International Journal of Applied Earth Observation and

Geoinformation **128** (2024)

[13] Zhou, F., Li, X.: Review of deep learning-based stereo matching. Displays **87**, 102940 (2025) https://doi.org/10.1016/j.displa.2024.102940

[14] Wu, B., Vallet, M., Pierrot-Deseilligny, M., *et al.*: An evaluation of deep learning based stereo dense matching dataset shift from aerial images and a large scale stereo dataset. International Journal of Applied Earth Observation and Geoinformation **128**, 103707 (2024) https://doi.org/10.1016/j.jag.2024.103707

[15] Zhao, Y., Liu, Y., Gao, S., Liu, G., Wan, Z., Hu, D.: Deep learning-based digital surface model reconstruction of zy-3 satellite imagery. Remote Sensing **16**(14), 2567 (2024) https://doi.org/10.3390/rs16142567

[16] Zhu, L., Hattula, E., Raninen, J., Hyyppä, J.: Experiment on producing disparity maps from aerial stereo images using unsupervised and supervised methods. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences **XLVIII-4/W1**, 561–568 (2022) https://doi.org/10.5194/isprs-archives-XLVIII-4-W1-2022-561-2022

[17] Madani, A., Author, B., Author, C.: Generative deep learning for high-resolution dem generation from stereo satellite imagery. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (2025). Early Access

[18] Albanwan, H., Qin, R.: Fine-tuning deep learning models for stereo matching using results from semi-global matching. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences **V-2**, 39–46 (2022) https://doi.org/10.5194/isprs-annals-V-2-2022-39-2022

[19] Tao, C.V.: A comprehensive study of the rational function model for photogrammetric processing. Photogrammetric Engineering & Remote Sensing **69**(4), 389–399 (2003)

[20] Grodecki, J., Dial, G.: Block adjustment of high-resolution satellite images described by rational polynomial coefficients. Photogrammetric Engineering and Remote Sensing **69**(1), 59–68 (2003)

[21] Chen, C., Zhao, L., Sun, X.: Semstereo: Semantic-constrained stereo matching network for remote sensing. arXiv preprint arXiv:2412.12685 (2024)