CONTINUOUS 3D-LABEL SEMI-GLOBAL MATCHING FOR SATELLITE STEREO

Sonali Patil[1,*] and Qi Guo[2]

[1] German Aerospace Center (DLR), Institute for Software Technology, Braunschweig, Germany - sonali.patil@dlr.de
[2] Purdue University, School of Electrical and Computer Engineering, West Lafayette, IN, USA - guo675@purdue.edu

KEY WORDS: Satellite Photogrammetry, Dense stereo matching, Stereo 3D Reconstruction, Digital Earth

ABSTRACT:

We present Continuous 3D-Label Semi-Global Matching (CoSGM), a new dense stereo matching algorithm for satellite stereo. CoSGM overparameterizes the disparity map as an array of local planes, so it can regularize the first-order smoothness of the estimated disparity. Furthermore, the algorithm can produce a dense normal map beside the disparity map. We show experimentally that CoSGM could achieve denser depth maps with comparable accuracy to traditional semi-global matching algorithms.

1. INTRODUCTION

Measuring three-dimensional terrain shapes is a critical problem in photogrammetry, as the estimated terrain shape could be broadly applied in environmental studies, national defense, navigation, etc. One of the emerging approaches nowadays to reconstructing terrain shapes is through stereo matching on satellite images; this is possible thanks to the increased popularity of very high resolution (VHR) optical imagers carried on satellites (Zhao et al., 2022), which can capture images of large terrain areas with resolution as high as 30cm per image pixel.

There exist satellite stereo approaches based on two-view stereo (Shean et al., 2016, d'Angelo and Kuschk, 2012, Rupnik et al., 2018, Qin, 2016, De Franchis et al., 2014, Youssefi et al., 2020) and multi-view stereo (Gómez et al., 2022, Ozcanli et al., 2015), and this paper concentrates on the two-view approaches. These methods first generate a density disparity map per each pair of VHR images, and then fuse the 3D information together to form a 3D representation of the entire terrain area, named the digital surface model (DSM).

Another unique characteristic of satellite stereo is that it is nontrivial to capture ground truth terrain shape (Patil and Guo, 2023), which imposes challenges to training deep stereo algorithms that require sufficient ground truth data. Currently, there are two major types of deep stereo algorithms. The first is based on the semi-global matching (SGM) pipeline that first generates a cost volume using deep networks, then performs cost aggregation to produce a disparity map (Zbontar et al., 2016, Zhang et al., 2019). The second is end-to-end, which directly uses deep architectures to output disparity maps from rectified image pairs (Lipson et al., 2021). A previous study has empirically shown that the SGM-based approaches are more robust and generalizable to unseen regions than end-to-end approaches (Albanwan and Qin, 2022a). This suggests that the SGM-based approaches are possibly more suitable for satellite stereo in the current stage, when the terrain types and areas in the ground truth data are limited. Albanwan *et al.* (Albanwan and Qin, 2022b) tried to address the generalizability issue of end-to-end architectures by using SGM as a teacher to finetune the model. This also advise the usefulness of SGM-based approaches for satellite stereo.

We present continuous 3D-label semi-global matching (CoSGM) for satellite stereo. It is a new cost aggregation algorithm for the SGM pipeline that uses an overparameterized strategy. Specifically, instead of parameterizing the disparity of each pixel as a single value, CoSGM parameterizes the local surface at each pixel as a plane, and the algorithm finds the optimal local plane for each pixel. This enables CoSGM to simultaneously output a dense normal map of the scene besides the disparity map (Fig. 1). We show that CoSGM has the same order of computational complexity as SGM and has a small memory overhead, but it can generate denser DSM with comparable accuracy with fewer stereo pairs when compared to using traditional cost aggregation strategies.

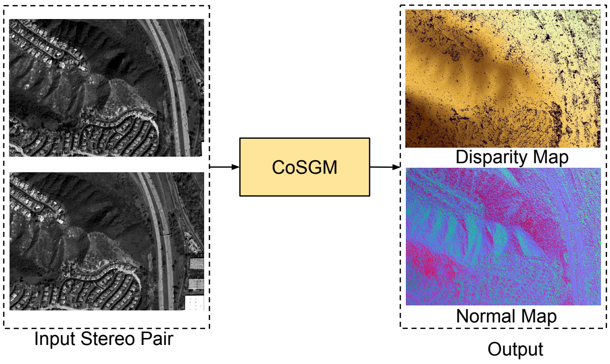


Figure 1. CoSGM simultaneously outputs disparity and normal maps. The plane parameterization overcomes front-parallel bias in SGM; the fronto-parallel bias is particularly apparent in undulating terrains. Our experimental results show CoSGM outperforms SGM in such regions. Additionally it produces dense disparity maps which helps to obtain 3D reconstruction with fewer stereo pairs than SGM.

2. BACKGROUND

SGM was originally derived from the earlier global energy minimizing function shown in Eq. 1. The objective is to estimate the disparity map that minimizes the given energy function, which is an NP-Hard problem (Boykov et al., 2001):

$$\min_{\mathbf{d}} \left\{ \sum_{\mathbf{p}} \left(\mathcal{C}(d_{\mathbf{p}}) + \lambda \sum_{\mathbf{q} \in \mathcal{N}_{\mathbf{p}}} \mathcal{V}(d_{\mathbf{p}}, d_{\mathbf{q}}) \right) \right\}, \tag{1}$$

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLVIII-1-W2-2023-1499-2023 | © Author(s) 2023. CC BY 4.0 License.

where $\mathbf{d} = \{d_\mathbf{p} | \forall \mathbf{p}\}$ is a list of disparities of all pixels $\mathbf{p}$. The cost $\mathcal{C}(d_\mathbf{p})$ is a unary data cost of assigning disparity $d_\mathbf{p} \in \{d_{min}, \cdots, d_{max}\}$ to pixel position $\mathbf{p} \in \mathbb{R}^2$. The cost $\mathcal{C}(d_\mathbf{p})$ for all pixels $\mathbf{p}$ and all possible disparities $d_\mathbf{p}$ is previously generated, either using traditional approaches, such as census transform, or learning-based approaches. Thus, the cost $\mathcal{C}(d_\mathbf{p})$ for all pixels $\mathbf{p}$ and all possible disparities $d_\mathbf{p}$ form a $H \times W \times D$ tensor $\mathcal{C}$, where H and W are the height and width of the disparity map, and D is the number of possible disparities. The pairwise smoothness $\mathcal{V}(d_\mathbf{p}, d_\mathbf{q})$ enforces that $d_\mathbf{p}$ is close to $d_\mathbf{q}$, where $\mathbf{q} \in \mathcal{N}_\mathbf{p}$ and $\mathcal{N}_\mathbf{p}$ is either a 4-connected or 8-connected neighborhood of pixel $\mathbf{p}$.

Instead, SGM solves the same problem along eight cardinal 1D directions $\mathbf{r} = \{(1,0), (-1,0), (0,1) \cdots\}$. The recursive minimization function $\mathcal{L}_\mathbf{r}$ per direction $\mathbf{r}$ is shown as follows:

$$\mathcal{L}_\mathbf{r}(d_\mathbf{p}) = \mathcal{C}(d_\mathbf{p}) + \min_{d' \in \Omega}(\mathcal{L}_\mathbf{r}(d_\mathbf{q}) + \mathcal{V}(d, d')) - \min_k \mathcal{L}_\mathbf{r}(k_\mathbf{q}), \tag{2}$$

where $\mathbf{q}$ is the previous pixel of $\mathbf{p}$ along direction $\mathbf{r}$, $\mathbf{q} = \mathbf{p} - \mathbf{r}$. The loss $\mathcal{L}_\mathbf{r}(d_\mathbf{q})$ is the cost of previous recursion along direction $\mathbf{r}$ at pixel $\mathbf{q}$ with disparity $d_\mathbf{q}$. Subtracting the term $\min_k \mathcal{L}_\mathbf{r}(k_\mathbf{q})$ controls the maximum value of the cost $\mathcal{L}_\mathbf{r}(d_\mathbf{q})$. It guarantees that $\mathcal{L}_\mathbf{r}(d_\mathbf{p}) \leq \mathcal{C}_{max} + P_2$, where $\mathcal{C}_{max}$ is the maximal value of the cost volume $\mathcal{C}$.

The pairwise or first-order smoothness term $\mathcal{V}(d, d')$ is given by Eq. 3, which favors front-parallel surfaces (Fig. 3(a).) The constant penalty assumes the local variation of disparity is near constant.

$$\mathcal{V}(d, d') = \begin{cases} 0 & \text{if } d = d' \\ P_1 & \text{if } |d - d'| = 1 \\ P_2 & \text{if } |d - d'| > 1 \end{cases} \tag{3}$$

If the neighboring pixel's disparity doesn't change then no penalty is assigned. Small penalty P_1 is assigned if the disparity varies by one pixel. This allows estimating slightly sloped or curved surfaces. Larger penalty $P_2 (P_2 > P_1)$ is assigned otherwise to preserve larger disparity discontinuities. P_1 and P_2 are input hyperparameters. These parameters are typically updated as a function of image intensity (Schonberger et al., 2018, Zbontar et al., 2016). People also tried different ways to define the smoothness term. For example, Zbontar *et al.* observed small changes in disparity to appear more frequently vertically than horizontally. Therefore, the penalty for disparity change in the vertical direction is lower than that of the horizontal direction.

The cost volumes $\mathcal{L}_\mathbf{r}$ obtained in each different direction $\mathbf{r}$ are summed to obtain a new cost volume:

$$\mathcal{S}(d_\mathbf{p}) = \sum_\mathbf{r} \mathcal{L}_\mathbf{r}(d_\mathbf{p}), \tag{4}$$

where $\mathcal{S}(d_\mathbf{p})$ indicates the cost for pixel $\mathbf{p}$ having disparity $d_\mathbf{p}$.

Drory *et al.* (Drory et al., 2014) showed the connection between SGM and non-loopy belief propagation on a special type of subgraph. They proposed overcounting correction as given in Eq. 5

$$\mathcal{S}(d_\mathbf{p}) = \mathcal{S}(d_\mathbf{p}) - 7\mathcal{C}(d_\mathbf{p}) \tag{5}$$

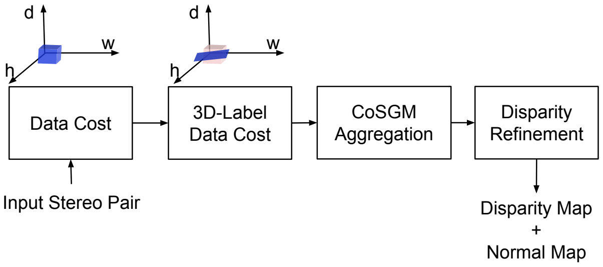


Figure 2. CoSGM algorithm. Given a stereo pair, first we compute the data cost similar to SGM and then we obtain 3D-label data cost (Eq. 10). CoSGM aggregation (Eqs. 12 and 13) produces disparity and normal maps. Normal map is used to further refine the disparity map.

This is due to the fact that the cost $\mathcal{C}(d_\mathbf{p})$ is added per direction $\mathcal{L}_\mathbf{r}(d_\mathbf{p})$. The overcount correction adjust this additional counting of cost before computing the disparity map.

The final disparity $\mathbf{D}(\mathbf{p})$ of each pixel $\mathbf{p}$ is computed by using a Winner-Takes-All (WTA) evaluation over the volume $\mathcal{S}(d_\mathbf{p})$.

$$\mathbf{D}(\mathbf{p}) = \arg\min_d \mathcal{S}(d_\mathbf{p}). \tag{6}$$

3. PROPOSED METHOD

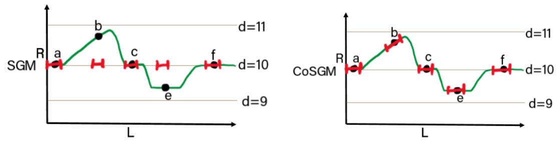
The CoSGM aims to solve a similar problem as Eq. 1:

$$\min_\mathbf{f} \left\{ \sum_\mathbf{p} \left(\phi(f_\mathbf{p}) + \lambda \sum_{\mathbf{q} \in \mathcal{N}_\mathbf{p}} \psi(f_\mathbf{p}, f_\mathbf{q}) \right) \right\}. \tag{7}$$

For each pixel $\mathbf{p}$, the vector $f_\mathbf{p} = (a, b, c)$ defines a local plane that fits the disparity at a neighbor $\mathcal{N}_\mathbf{p}$ centered at $\mathbf{p}$. For a point $\mathbf{q} \in \mathcal{N}_\mathbf{p}$, the plane $f_\mathbf{p}$ gives the disparity at $\mathbf{q}$ as:

$$d_\mathbf{q}(f_\mathbf{p}) = [a, b, c][q_x, q_y, 1]^T, \tag{8}$$

where $\mathbf{q} = [q_x, q_y]$. The first term in Eq. 7, $\phi(f_\mathbf{p})$, is a unary data cost of the local plane $f_\mathbf{p}$ defined at pixel $\mathbf{p}$, and $\psi(f_\mathbf{p}, f_\mathbf{q})$ is pairwise smoothness term between the local planes $f_\mathbf{p}, f_\mathbf{q}$ centered at pixels $\mathbf{p}$ and $\mathbf{q}$. The term f indicates the list of local planes for all pixels $\mathbf{f} = \{f_\mathbf{p} | \forall \mathbf{p}\}$. This overparameterization of disparity enables the optimization to exert first-order smoothness regularization to the reconstructed disparity map (Fig. 3).



(a) Fronto-parallel bias in SGM

(b) CoSGM Motivation

Figure 3. The zeroth order smoothness in SGM penalizes the disparity value among neighboring pixels to be different from each other. Thus, it could cause fronto parallel bias and makes it challenging in regions with undulating terrain. CoSGM uses overparameterized disparity planes to enable first-order smoothness, which regularizes the first-order derivative of the disparity map to be smooth.

3.1 3D-Label Data Cost

The overparameterization $f_{\mathbf{p}}$ introduces three unknowns a, b, c to solve for each pixel $\mathbf{p}$. To reduce computational complexity, we constrain the possible planes $f_{\mathbf{p}}$ for each pixel $\mathbf{p}$ to a finite list of candidates $\{f^d_{\mathbf{p}} | d = d_{min}, \cdots, d_{max}\}$.

Fig. 4 demonstrates how each local plane candidate $f^d_{\mathbf{p}}$ is calculated. For each pixel $\mathbf{p}$ and each potential disparity value for this pixel d, we iterate every pixel $\mathbf{p}_i$ in the neighborhood $\mathcal{N}_{\mathbf{p}}$. For each pixel $\mathbf{p}_i$, we find the optimal local disparity d_i according to:

$$d_i = \operatorname*{arg\,min}_{d'_{\mathbf{p}_i} \in \{d-1, d, d+1\}} \mathcal{C}(d'_{\mathbf{p}_i}). \tag{9}$$

Then, we can build a linear system that fits a plane to the optimal local disparities d_i of all pixels $\mathbf{p}_i$ in the neighborhood $\mathcal{N}_{\mathbf{p}}$:

$$\begin{bmatrix} p^x_1 & p^y_1 & 1 \\ p^x_2 & p^y_2 & 1 \\ \vdots & \vdots & \vdots \\ p^x_i & p^y_i & 1 \\ \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_i \\ \vdots \end{bmatrix}. \tag{10}$$

The corresponding plane is the solution of this linear system $f^d_{\mathbf{p}} = [a, b, c]$.

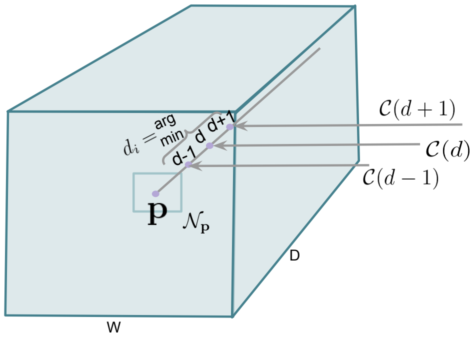


Figure 4. The process of finding d_i for Eq. 10 to solve the local plane candidate $f^d_{\mathbf{p}}$.

For each local plane candidate $f^d_{\mathbf{p}}$, the disparity at $\mathbf{p}$ is not necessarily an integer. We define the unary cost at $\mathbf{p}$ for plane $f^d_{\mathbf{p}}$ by linearly interpolate cost volume $\mathcal{C}(d_{\mathbf{p}})$ as:

$$\begin{aligned} \phi(f^d_{\mathbf{p}}) &= (d - \lfloor d_{\mathbf{p}}(f^d_{\mathbf{p}}) \rfloor) \mathcal{C}(\lfloor d_{\mathbf{p}}(f^d_{\mathbf{p}}) \rfloor + 1) \\ &+ (\lfloor d_{\mathbf{p}}(f^d_{\mathbf{p}}) \rfloor + 1 - d) \mathcal{C}(\lfloor d_{\mathbf{p}}(f^d_{\mathbf{p}}) \rfloor). \end{aligned} \tag{11}$$

3.2 CoSGM Aggregation

Eq. 12 shows our CoSGM aggregation formulation:

$$\begin{aligned} \mathcal{L}_{\mathbf{r}}(f^d_{\mathbf{p}}) = \phi(f^d_{\mathbf{p}}) &+ \min_{d' \in \Omega} (\mathcal{L}_{\mathbf{r}}(f^d_{\mathbf{q}}) + \mathcal{V}(f^d_{\mathbf{p}}, f^{d'}_{\mathbf{q}})) \\ &- \min_{k \in \Omega} \mathcal{L}_{\mathbf{r}}(f^k_{\mathbf{q}}) \end{aligned} \tag{12}$$

where $\mathcal{V}(f^d_{\mathbf{p}}, f^{d'}_{\mathbf{q}})$ is the smoothness term between local plane candidates $f^d_{\mathbf{p}}, f^{d'}_{\mathbf{q}}$ of adjacent pixels $\mathbf{p}, \mathbf{q}$. We define the smoothness term as follows:

$$\mathcal{V}(f^d_{\mathbf{p}}, f^{d'}_{\mathbf{q}}) = \begin{cases} 0 & \text{if } d = d' \\ \psi(f^d_{\mathbf{p}}, f^{d'}_{\mathbf{q}}) & \text{if } d \neq d' \end{cases} \tag{13}$$

where

$$\psi(f^d_{\mathbf{p}}, f^{d'}_{\mathbf{q}}) = \begin{cases} \alpha_1 \cdot \max(w_{\mathbf{p},\mathbf{q}}, \epsilon) \cdot \min(\overline{\psi}(f^d_{\mathbf{p}}, f^{d'}_{\mathbf{q}}), \tau) \\ \qquad \text{if } |d - d'| = 1 \\ \alpha_2 \cdot \max(w_{\mathbf{p},\mathbf{q}}, \epsilon) \cdot \min(\overline{\psi}(f^d_{\mathbf{p}}, f^{d'}_{\mathbf{q}}), \tau) \\ \qquad \text{if } |d - d'| > 1. \end{cases} \tag{14}$$

The weight $w_{\mathbf{p},\mathbf{q}} = e^{-||\mathbf{I}_L(\mathbf{p}) - \mathbf{I}_L(\mathbf{q})||_1/\gamma}$, where $\mathbf{I}_L$ is the left image in a given stereo pair. The term $\overline{\psi}(f^d_{\mathbf{p}}, f^{d'}_{\mathbf{q}})$ is the difference between two adjacent local planes defined below and illustrated in Fig. 5:

$$\overline{\psi}(f^d_{\mathbf{p}}, f^{d'}_{\mathbf{q}}) = |d_{\mathbf{p}}(f^d_{\mathbf{p}}) - d_{\mathbf{p}}(f^{d'}_{\mathbf{q}})| + |d_{\mathbf{q}}(f^{d'}_{\mathbf{q}}) - d_{\mathbf{q}}(f^d_{\mathbf{p}})| \tag{15}$$

We obtain $\psi(f^d_{\mathbf{p}}, f^{d'}_{\mathbf{q}})$ from $\overline{\psi}(f^d_{\mathbf{p}}, f^{d'}_{\mathbf{q}})$ by truncating the distance by the hyperparameters τ. Multiplying with $\max(w_{\mathbf{p},\mathbf{q}}, \epsilon)$ helps to guide the aggregation by image intensity information.

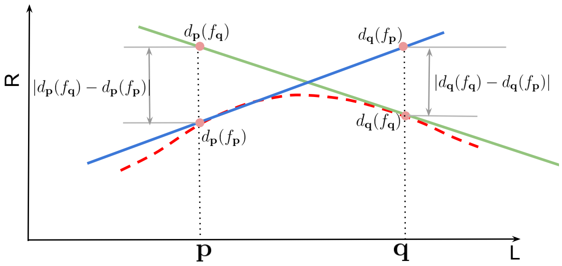


Figure 5. Smoothness cost defined over disparity planes. The x-axis represent pixel positions in the left image (L) and the y-axis represent pixel positions in the corresponding right image (R) along the same row as (L).

For pixels in the horizontal paths, (α_1, α_2) is changed using the following formula.

$$(\alpha_1, \alpha_2) = \begin{cases} (\alpha_1, \alpha_2) & \text{if } \; D_1 \leq \beta \text{ and } D_2 \leq \beta \\ (\alpha_1/Q_2, \alpha_2/Q_2) & \text{if } \; D_1 \geq \beta \text{ and } D_2 \geq \beta \\ (\alpha_1/Q_1, \alpha_2/Q_1) & \quad \text{otherwise} \end{cases} \tag{16}$$

For pixels in the vertical paths, following the recommendation made in (Zbontar et al., 2016), after updating (α_1, α_2) using Eq. 16, we set

$$\alpha_1 = \alpha_1 / V \tag{17}$$

And, when a pixel is on a diagonal path in the support structure, we set

$$\alpha_1 = \alpha_1 * \frac{\sqrt{1.0 + V^2}}{V} \tag{18}$$

We sum all the CoSGM 1D aggregation cost volumes $\mathcal{L}_{\mathbf{r}}(f^d_{\mathbf{p}})$.

$$\mathcal{S}(f^d_{\mathbf{p}}) = \sum_{\mathbf{r}} \mathcal{L}_{\mathbf{r}}(f^d_{\mathbf{p}}) \tag{19}$$

We apply overcounting correction before taking WTA.

$$\mathcal{S}(f_{\mathbf{p}}^d) = \mathcal{S}(f_{\mathbf{p}}^d) - 7\phi(f_{\mathbf{p}}^d) \tag{20}$$

Furthermore, we can optionally produce normal map using the disparity plane parameters (a, b, c) as $n_z = 1/(1 + a^2 + b^2)$, $n_x = -an_z$ and $n_y = -bn_z$.

Some interesting similarities with SGM aggregation are as follows.

- Subtracting $\min_{k\in\Omega} \mathcal{L}_{\mathbf{r}}(f_{\mathbf{q}}^k)$ ensures that $\mathcal{L}_{\mathbf{r}} \leq \phi_{max} + \alpha_2 \cdot \tau$
- Setting $\alpha_2 >> \alpha_1$ ensures that disparity levels greater than one from the current disparity d penalized higher.
- SGM overcount correction given by (Drory et al., 2014) can be easily adapted for CoSGM.

3.3 Disparity Refinement

Dense stereo matching between $\mathbf{I}_L$ and $\mathbf{I}_R$ produces a disparity map $\mathbf{D}'_L$ with respect to the left image. We obtain $\mathbf{D}'_R$ by swapping the order of the input stereo pair. Then we can perform a consistency check to enforce the uniqueness constraint on $\mathbf{D}'_L$ and $\mathbf{D}'_R$ as described in Patil and Guo (Patil and Guo, 2023). The consistency check would produce new disparity maps $\mathbf{D}_L$ and $\mathbf{D}_R$ with pixels marked invalid wherever the uniqueness constraint failed. We further refine disparity maps using the Joint Bilateral Filter (Yang et al., 2010). It refines the disparity map $\mathbf{D}_L$ while preserving the edges in $\mathbf{I}_L$.

3.4 Hierarchical Optimization

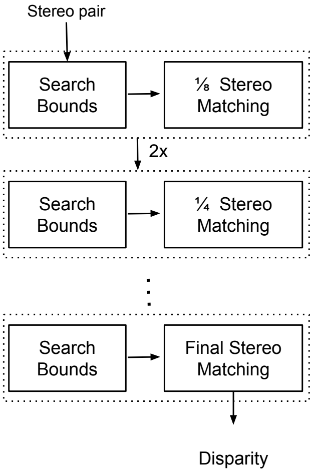


Figure 6. We process large stereo pairs hierarchically and use disparity maps from lower resolution to estimate disparity bounds for the next level of resolution.

Rothermal *et al.* (Rothermel et al., 2012) proposed tSGM, which is a hierarchical version of SGM. The tSGM processes a stereo pair at a lower resolution, performs the consistency check, and uses the low-resolution disparity map as a guide for the next higher resolution. They showed that not only does tSGM produce disparity maps faster than SGM, but the output disparity map is also denser. Patil *et al.* propose a modification to tSGM by estimating the per-pixel search bound by *DEM-Sculpting* (Patil et al., 2019). Fig. 6 shows key steps of their modified tSGM. We use a similar approach in our CoSGM to estimate large disparity maps efficiently.

4. PRELIMINARY EXPERIMENTS

In this section, we provide some preliminary comparisons of CoSGM with other satellite stereo methods. We analyze the quality of the 3D terrain reconstruction in the format of digital surface models (DSMs), which is a raster with georeferenced height at each point. We use the DSM generation pipeline of Stellar (Patil and Guo, 2023) to produce the DSMs (Fig. 7). The input to the pipeline consists of a set of multi-date and multi-view VHR satellite images and their corresponding camera parameters represented as Rational Polynomial Coefficients (RPC) (Baltsavias and Stallmann, 1992). The images and camera parameters are then preprocessed to generate stereo image pairs, which include camera calibration and stereo pair selection. Then, the pipeline performs stereo matching to generate disparity maps and transforms these disparity maps into pairwise DSMs. Finally, the pipeline aggregates the pairwise DSMs together to generate a single DSM for the area.

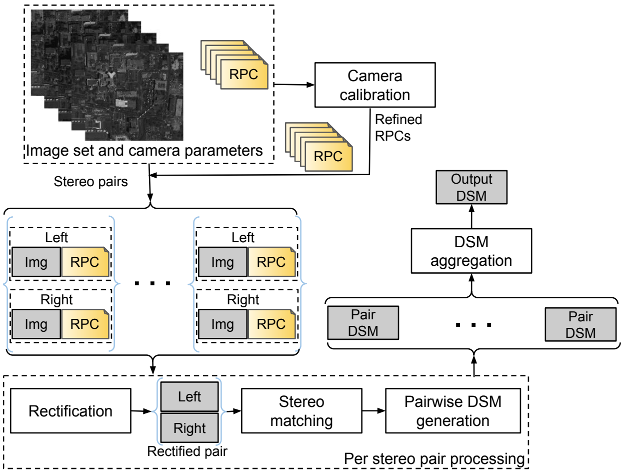


Figure 7. Overview of DSM generation from multiple satellite stereo pairs.

Preprocessing and Stereo Rectification Given a set of input images and the corresponding Rational Polynomial Coefficients (RPC) camera models. We first calibrate the camera parameters using Bundle adjustment. Then, we perform rectification for each stereo pair. The process involves first stereo rectifying 500 × 500 tiles with affine camera model assumption, then using Digital Elevation Model (DEM) from Shuttle Radar Topography Mission (SRTM) with 30 meters resolution to find virtual correspondences for estimating stereo rectification tomography, and finally producing a large stereo rectified images by stitching small stereo rectified tiles together. We use the homography to further obtain stereo rectification which produces unipolar disparity maps. For details, refer to Patil and Guo (Patil and Guo, 2023).

In this section, we provide two analyses of CoSGM. First, we compare pairwise DSMs generated from CoSGM with those from the modified tSGM (Patil et al., 2019) on the Stellar dataset in Sec. 4.1. Then in Sec. 4.2, we run CoSGM on a subset of the IARPA Challenge dataset, and generate an aggregated DSM following the procedure in Fig. 7. We compare the numbers with two top-performing DSM generation pipelines reported by the IARPA benchmark (Bosch et al., 2017, Michel et al., 2020). We use the following hyperparameters for SGM and CoSGM. For SGM we set the hyperparameters as follows $P_1 = 24$, $P_2 = 144$, $Q_1 = 4$, $Q_2 = 8$, $V = 1.4$, $\beta = 2$ and for CoSGM $\alpha_1 = 50$, $\alpha_2 = 300$, $\epsilon = 0.03$, $\tau = 20$ and $\gamma = 10$.

We use the following evaluation metrics for both experiments.

- **Completeness (Comp.)**: Percentage of points in the estimated DSM where the absolute altitude difference is less than 1 meter with respect to the ground truth lidar DSM. The higher number indicates the denser stereo DSM.

- **RMSE**: Root Mean Square Error (RMSE) over valid pixels in both estimated and ground truth DSM. Lower RMSE is preferred in the estimated DSM.

- **MAE**: Median Absolute Z-Error (MAE) over valid pixels in both DSMs. Lower MAE indicated more accurate reconstruction

4.1 Stellar Dataset Results

Recently Patil *et al.* (Patil and Guo, 2023) created a very large dataset for DSM evaluation from multi-date satellite images. It includes ground truth Lidar DSM for five cities in North and South America spanning more than 1500 km^2 and WorldView-3 images with 30 cm resolution. They additionally provide stereo-rectified pairs for a subset. The stereo pairs cover approximately 2 km^2 regions. The stereo-rectified images are approximately 5000 × 5000. We chose three regions, including San Diego, California, Omaha, Nebraska, and Jacksonville. San Diego has undulating terrain, and the regions in Jacksonville and Omaha are relatively flatter. Jacksonville also has a water body. Therefore, these regions are well representative of reconstruction challenges due to topography changes. We processed 103 stereo pairs in the region in Jacksonville, 273 in the region in Omaha, and 253 in the region in San Diego. Table 1 shows average Completeness, MAE, and RMSE for each region. For San Diego, the average completeness score is higher than the modified tSGM, and MAE and RMSE are comparable. This indicates denser reconstruction without compromising accuracy too much.

Region		Comp (%) ↑	MAE (m) ↓	RMSE (m) ↓
M. tSGM	San Diego Omaha Jacksonville	30.7 28.8 30.7	0.89 1.41 0.89	3.19 9.09 7.24
CoSGM	San Diego Omaha Jacksonville	32.1 29.8 27.0	0.94 1.54 0.73	3.68 10.23 8.07

Table 1. Quantiative evaluation of pairwise DSMs on Stellar dataset for the proposed CoSGM and the modified tSGM (Patil et al., 2019).

On average, CoSGM produces denser DSM in San Diego and comparable results in relatively flatter Omaha and Jacksonville. Jacksonville has a water region, and we haven't masked the water area before evaluating the results. This could explain the anomaly in the average results. Fig. 8 shows sample stereo DSMs from each region. In the San Diego region, the left image is from 22 November 2014, and the right image is from 24 February 2015. In the Omaha region, the left image is from 04 October 2014, and the right image is from 22 October 2014. In the Jacksonville region, the left image is from 15 February 2015, and the right image is from 27 January 2015. All these stereo pairs show improved results over tSGM. SGM's sensitivity to acquisition time difference (De Franchis et al., 2014), baseline (Carl et al., 2013, d'Angelo et al., 2014, Zhu et al., 2008), and sun angle difference (Qin, 2019) between stereo pairs is well studied. However, being a new algorithm, CoSGM's sensitivity to these parameters requires further future investigation.

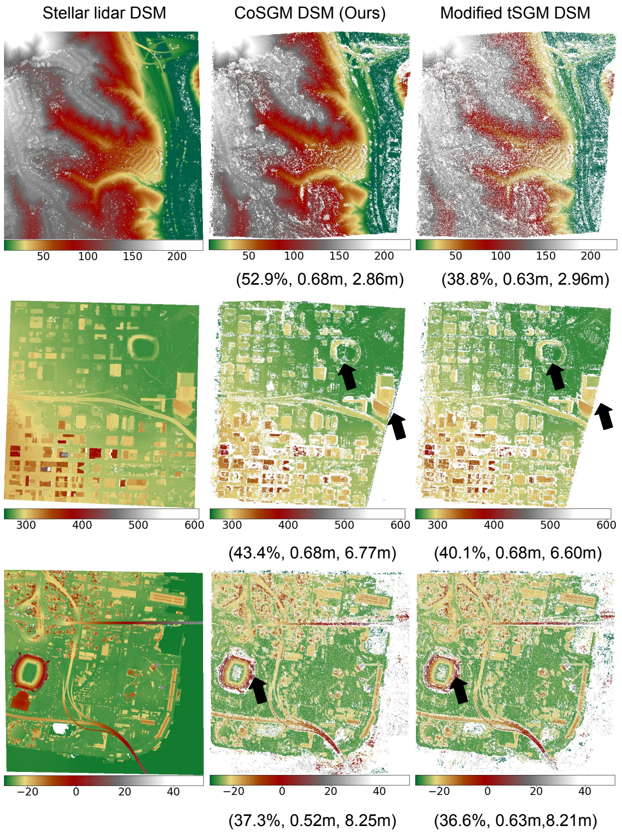


Figure 8. Stereo DSM results. From top to bottom: San Diego, Omaha and Jacksonville. The numbers are Completeness, MAE, and RMSE, respectively. For undulating terrain in San Diego, the improvement in density by CoSGM is significant over the modified tSGM.

4.2 IARPA Challenge Data Subset Results

In this experiment, we analyze the aggregated DSMs of a large area generated by CoSGM, and compare them with two top-performing DSM generation pipelines of the IARPA benchmark, S2P (De Franchis et al., 2014) and CARS (Michel et al., 2020). Both S2P and CARS are based on variants of the SGM. S2P uses the More Global Matching (MGM) algorithm (Facciolo et al., 2015) for stereo matching, and CARS uses the vanilla SGM. This experiment uses six WorldView-3 in-track images acquired on 18 December 2015 from the IARPA dataset. Table 2 shows the quantitative comparison results. Each method first uses a certain number of stereo pairs to generate disparity maps, and then aggregates the disparity maps together to produce a single DSM. The number of stereo pairs used are listed in the table. CoSGM achieves comparable performance to S2P (De Franchis et al., 2014) using only five stereo pairs and get denser DSM without compromising accuracy much with 15 stereo pairs, while S2P requires 50 pairs of stereo images. The results of CoSGM and CARS are complementary; CoSGM achieves denser DSMs while having higher mean errors. Fig. 9 shows qualitative results of DSMs obtained by aggregating five stereo pairs and 15 stereo pairs using CoSGM. It demonstrates how the DSM quality improves by using more stereo pairs.

Method	Comp. (%) ↑	MAE $(m)\downarrow$	RMSE $(m)\downarrow$
CoSGM (5 stereo pairs)	73	0.35	2.59
CoSGM (15 stereo pairs)	75.6	0.26	2.65
S2P (50 stereo pairs)	73.2	0.37	2.59
CARS (4 stereo pairs)	68.39	0.24	2.19

Table 2. Fused DSM comparison with the S2P (De Franchis et al., 2014) and CARS pipelines (Michel et al., 2020). CoSGM achieves comparable accuracy to and higher density than S2P using much fewer stereo pairs.

REFERENCES

Albanwan, H., Qin, R., 2022a. A Comparative Study on Deep-Learning Methods for Dense Image Matching of Multi-angle and Multi-date Remote Sensing Stereo Images. *arXiv preprint arXiv:2210.14031*.

Albanwan, H., Qin, R., 2022b. Fine-tuning deep learning models for stereo matching using results from semi-global matching. *arXiv preprint arXiv:2205.14051*.

Baltsavias, E. P., Stallmann, D., 1992. Metric information extraction from spot images and the role of polynomial mapping functions. *XVII ISPRS Congress, Commission IV*, Swiss Federal Institute of Technology, Institute of Geodesy and Photogrammetry.

Bosch, M., Leichtman, A., Chilcott, D., Goldberg, H., Brown, M., 2017. METRIC EVALUATION PIPELINE FOR 3D MODELING OF URBAN SCENES. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 42.

Boykov, Y., Veksler, O., Zabih, R., 2001. Fast approximate energy minimization via graph cuts. *IEEE Transactions on pattern analysis and machine intelligence*, 23(11), 1222–1239.

Carl, S., Bärisch, S., Lang, F., d'Angelo, P., Arefi, H., Reinartz, P., 2013. Operational Generation of High-Resolution Digital Surface Models from Commercial Tri-Stereo Satellite Data. *Photogrammetric Week'13*, 261–269.

d'Angelo, P., Kuschk, G., 2012. Dense multi-view stereo from satellite imagery. *2012 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, 6944–6947.

d'Angelo, P., Rossi, C., Minet, C., Eineder, M., Flory, M., Niemeyer, I., 2014. High resolution 3d earth observation data analysis for safeguards activities. *Symposium on International Safeguards*, 1–8.

De Franchis, C., Meinhardt-Llopis, E., Michel, J., Morel, J.-M., Facciolo, G., 2014. An automatic and modular stereo pipeline for pushbroom images. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*.

Drory, A., Haubold, C., Avidan, S., Hamprecht, F. A., 2014. Semi-global matching: a principled derivation in terms of message passing. *German Conference on Pattern Recognition*, Springer, 43–53.

Facciolo, G., De Franchis, C., Meinhardt, E., 2015. Mgm: A significantly more global matching for stereovision. *BMVC 2015*.

Gómez, A., Randall, G., Facciolo, G., von Gioi, R. G., 2022. An experimental comparison of multi-view stereo approaches on satellite images. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 844–853.

Lipson, L., Teed, Z., Deng, J., 2021. Raft-stereo: Multilevel recurrent field transforms for stereo matching. *2021 International Conference on 3D Vision (3DV)*, IEEE, 218–227.

Michel, J., Sarrazin, E., Youssefi, D., Cournet, M., Buffe, F., Delvit, J., Emilien, A., Bosman, J., Melet, O., L'Helguen, C., 2020. A new satellite imagery stereo pipeline designed for scalability, robustness and performance. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2, 171–178.

Ozcanli, O. C., Dong, Y., Mundy, J. L., Webb, H., Hammoud, R., Tom, V., 2015. A comparison of stereo and multiview 3-d reconstruction using cross-sensor satellite imagery. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 17–25.

Patil, S., 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, 433–440. https://isprs-archives.copernicus.org/articles/XLVIII-M-1-2023/433/2023/.

Patil, S., Prakash, T., Comandur, B., Kak, A., 2019. A comparative evaluation of SGM variants (including a new variant, tMGM) for dense stereo matching. *arXiv preprint arXiv:1911.09800*.

Qin, R., 2016. Rpc stereo processor (rsp)–a software package for digital surface model and orthophoto generation from satellite stereo imagery. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 3, 77–82.

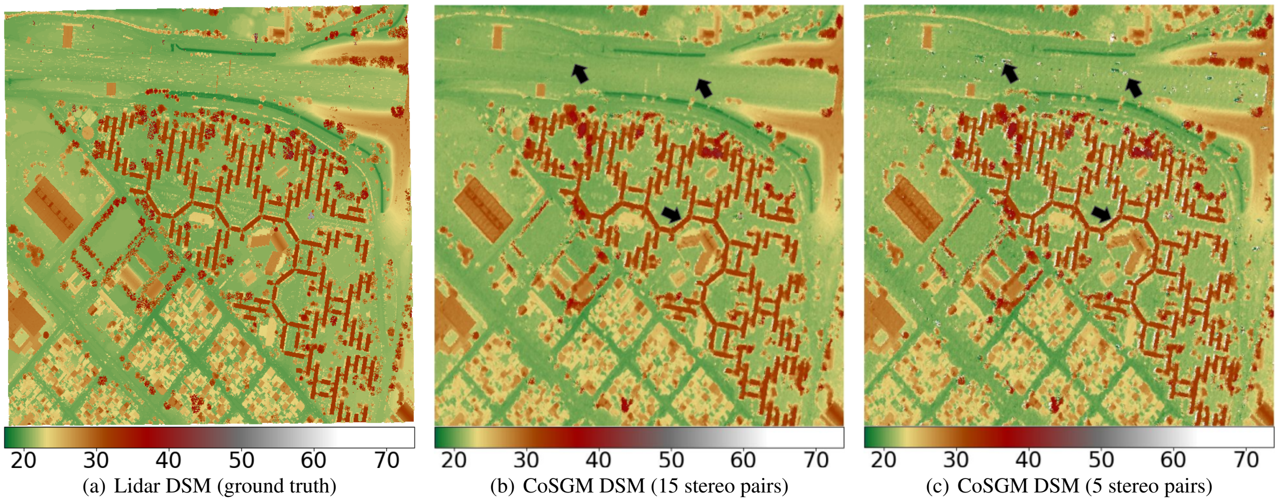
Qin, R., 2019. A critical analysis of satellite stereo pairs for digital surface model generation and a matching quality prediction model. *ISPRS Journal of Photogrammetry and Remote Sensing*, 154, 139–150.

Rothermel, M., Wenzel, K., Fritsch, D., Haala, N., 2012. Sure: Photogrammetric surface reconstruction from imagery. *Proceedings LC3D Workshop, Berlin*, 8number 2.

Rupnik, E., Pierrot-Deseilligny, M., Delorme, A., 2018. 3D reconstruction from multi-view VHR-satellite images in MicMac. *ISPRS Journal of Photogrammetry and Remote Sensing*, 139, 201–211.

Schonberger, J. L., Sinha, S. N., Pollefeys, M., 2018. Learning to fuse proposals from multiple scanline optimizations in semi-global matching. *Proceedings of the European Conference on Computer Vision (ECCV)*, 739–755.

Shean, D. E., Alexandrov, O., Moratto, Z. M., Smith, B. E., Joughin, I. R., Porter, C., Morin, P., 2016. An automated, open-source pipeline for mass production of digital elevation models (DEMs) from very-high-resolution commercial stereo satellite imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 116, 101–117.



(a) Lidar DSM (ground truth)

(b) CoSGM DSM (15 stereo pairs)

(c) CoSGM DSM (5 stereo pairs)

Figure 9. Qualitative results for fused DSMs using five pairs and 15 pairs. Arrows highlight improvements when using more stereo pairs.

Yang, Q., Wang, L., Ahuja, N., 2010. A constant-space belief propagation algorithm for stereo matching. *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE, 1458–1465.

Youssefi, D., Michel, J., Sarrazin, E., Buffe, F., Cournet, M., Delvit, J.-M., L'Helguen, C., Melet, O., Emilien, A., Bosman, J., 2020. Cars: A photogrammetry pipeline using dask graphs to construct a global 3d model. *IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, 453–456.

Zbontar, J., LeCun, Y. et al., 2016. Stereo Matching by Training a Convolutional Neural Network to Compare Image Patches. *Journal of Machine Learning Research*, 17(1-32), 2.

Zhang, K., Snavely, N., Sun, J., 2019. Leveraging vision reconstruction pipelines for satellite imagery. *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, 0–0.

Zhao, Q., Yu, L., Du, Z., Peng, D., Hao, P., Zhang, Y., Gong, P., 2022. An overview of the applications of earth observation satellite data: impacts and future trends. *Remote Sensing*, 14(8), 1863.

Zhu, L., Umakawa, H., Guan, F., Tachibana, K., Shimamura, H., 2008. Accuracy investigation of orthoimages obtained from high resolution satellite stereo pairs. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 37, 1145–1148.