Edge-Preserving Stereo Matching Using LiDAR Points and Image Line Features

Siyuan Zou, Xinyi Liu, Xu Huang, Yongjun Zhang, *Member, IEEE*,
Senyuan Wang, Shuang Wu, Zhi Zheng, and Bingxin Liu

Abstract— This letter proposes a LiDAR and image line-guided stereo matching method (L2GSM), which combines sparse but high-accuracy LiDAR points and sharp object edges of images to generate accurate and fine-structure point clouds. After extracting depth discontinuity lines on the image by using LiDAR depth information, we propose a trilateral update of cost volume and depth discontinuity lines-aware semi-global matching (SGM) strategies to integrate LiDAR data and depth discontinuity lines into the dense matching algorithm. The experimental results for the indoor and aerial datasets show that our method significantly improves the results of the original SGM and outperforms two state-of-the-art LiDAR constraints' SGM methods, especially in recovering the 3-D structure of low-textured and depth discontinuity regions. In addition, the 3-D point clouds generated by our proposed method outperform the LiDAR data and dense matching point clouds generated by Metashape and SURE aerial in terms of completeness and edge accuracy.

Index Terms— 3-D reconstruction, depth discontinuity lines, LiDAR, low-textured, semi-global matching (SGM), stereo matching.

I. INTRODUCTION

DENSE stereo matching, finding the disparity between matching pixels on a stereo image pair [1], has many influential applications in photogrammetry and computer vision [2]. However, conventional dense matching methods have weak distinguishability in low-textured and depth dis-continuity regions due to texture sensitivity [3], [4], as shown in Fig. 1(c). By contrast, LiDAR point clouds have high geo-metric accuracy independent of feature spectra [5]. However, as shown in Fig. 1(a) and (b), unlike pixel measurements from cameras, LiDAR data are sparse in most cases, which can lead to depth discontinuity regions that are not well reconstructed [6]. Therefore, the complementary fusion of LiDAR data and image data is a promising solution for generating accurate and fine-structure 3-D point clouds [7].

Dense image matching (DIM) constrained by LiDAR data, which integrates the LiDAR data into an advanced DIM framework [8], is a reliable method to fuse LiDAR data and images. LiDAR data can be used in many aspects of DIM, such as reducing the disparity search range [9] and optimizing the matching cost [10]. The aforementioned methods produce better-matching results than the original DIM algorithm but tend to generate extended regions at the depth discontinuity edges due to the smoothness constraint of dense matching. Reducing penalty constraints on gradient [11], canny [12], or texture [13] edges can improve the matching results in edge regions. However, the aforementioned strategy does not distinguish texture edges from geometric edges, which conflicts with penalty constraints tending to be changed at geometric edges. With the introduction of LiDAR data, it is feasible to distinguish texture edges from geometric edges. Hence, an edge-preserving stereo matching with LiDAR data and geometric edge knowledge is proposed to recover accurate 3-D structures of low-textured and depth discontinuity regions, as shown in Fig. 1(d).

II. METHODOLOGY

The inputs of the proposed method are stereo images and LiDAR point clouds located in the same area. The accurate registration [14] is a prerequisite for most of the LiDAR and image fusion methods including our proposed method. In preprocessing, stereo images need to be rectified in the epipolar image space, and LiDAR outliers need to be filtered out via positional uncertainty [15]. First, we triangulate a mesh surface from the LiDAR point clouds and generate an initial disparity map. Next, we extract depth discontinuity lines, which are located in the depth discontinuity regions, from the image line features based on the initial disparity map. Finally, we propose a trilateral update of cost volume for improving low-textured regions and depth discontinuity lines-aware semi-global matching (SGM) for preserving the depth discontinuity.

A. Depth Discontinuity Line Extraction

We project the LiDAR mesh surface onto the stereo epipolar images to generate an initial disparity map and LiDAR projec-tion points. The LiDAR projection points render sparse laser points with known depths or disparities in an image grid. The occluded LiDAR points are excluded when back-projecting the watertight LiDAR mesh to the image.

Straight lines provide important information on man-made objects, which are often within the boundaries of objects. Per the literature [16], depth discontinuity lines can be clearly identified by the initial disparity map. As shown in Fig. 2, two buffer rectangles are symmetrical to the straight line in the basic image with its length parallel to the straight line
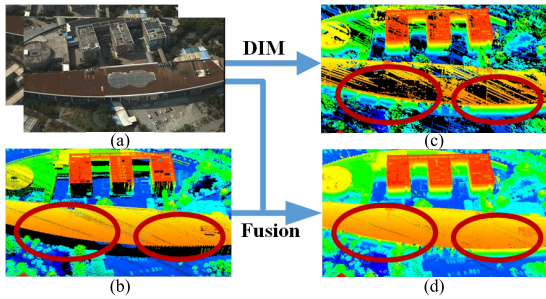
Fig. 1. Complementary fusion of images and LiDAR data. (a) Images. (b) LiDAR data. (c) DIM. (d) Proposed method (DIM constrained by LiDAR data).
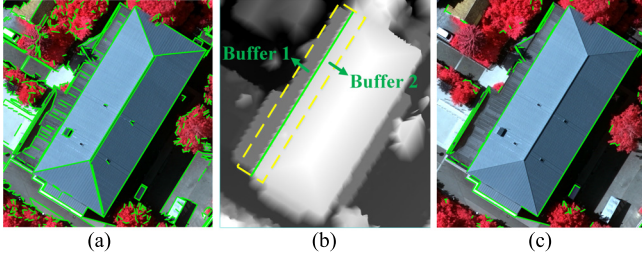

Fig. 2. Depth discontinuity line extraction. (a) Lines detected by LSD. (b) Initial disparity map. (c) Depth discontinuity lines.

and its width (ten pixels) perpendicular to the straight line. The median disparity value of the two buffer rectangles of the initial disparity map can be calculated. Only the straight line with obvious disparity changes (>one pixel) is selected as depth discontinuity lines.

B. Trilateral Update of Cost Volume

Cost volume stores the matching cost of each pixel in the reference image with the matching image in the disparity range. Cost volume optimization for updating the matching cost of homogeneous pixels when LiDAR data are available is an effective strategy for DIM constrained by LiDAR data [7]. Homogeneous pixels are pixels with no significant depth changes within the spatial neighborhood of the LiDAR projection point. Bilateral filtering based on the spatial and intensity domains is a commonly used method to determine homogeneous pixels, which defines a similarity measure between the central pixel q and each surrounding pixel p. The equation for calculating the similarity measure is as follows:

$$W_{(p,q)} = f(||p - q||)g(||I_p - I_q||) \tag{1}$$

where I_p and I_q are the intensities of the current and the centric pixel, respectively. The weighting functions, $f(\cdot)$ and $g(\cdot)$ are based on a Gaussian distribution. Whether p is a homogeneous pixel of q can be determined by applying a fixed threshold truncation (the empirical value is 0.7) to $W_{(p,q)}$. An illustration of the homogeneous pixels (yellow points) based on a bilateral filter is shown in Fig. 3(a). The right part of the homogeneous pixels is not in the continuous disparity region with the LiDAR projection point (red point). Thus, the aforementioned results conflict with the basic assumption that disparities of homogeneous pixels should be continuously varying.

By using the a priori of the depth discontinuity lines, a new solution to the aforementioned problem is available. We added depth discontinuity lines to the homogeneous pixel similarity metric to enhance the possibility that homogeneous pixels lay
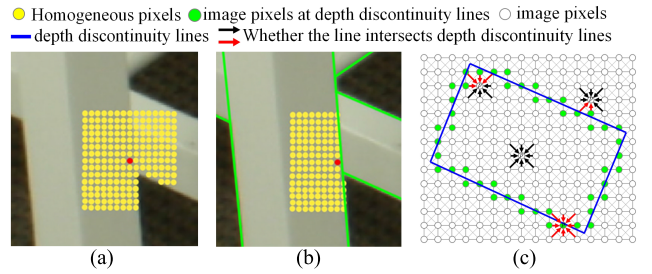

Fig. 3. Cost volume and cost aggregation optimization. (a) Bilateral update. (b) Trilateral update. (c) Depth discontinuity lines-aware SGM.

in regions of continuous disparity variation. We simplified the trilateral filter on the image space. First, we stored each depth discontinuity line in the image raster in the form of four neighborhoods. Next, after two pixels satisfied the bilateral filtering, each pixel on the line segment formed by these two pixels that did not contain depth discontinuity lines was considered a homogeneous pixel

$$W_{(p,q)} = f(||p - q||)g(||I_p - I_q||)T\{M \cap \mathrm{DL} = \varnothing\} \tag{2}$$

where $T\{\cdot\}$ is an indicator function that tests whether the line between the center pixel and the current pixel crosses the depth discontinuity lines, M denotes the binarized set of the line between the center pixel and the current pixel in the image space, and DL denotes the binarized set of depth discontinuity lines in the image space

$$T\{M \cap \mathrm{DL} = \varnothing\} = \begin{cases} 1 & M \cap \mathrm{DL} = \varnothing \\ 0 & M \cap \mathrm{DL} \neq \varnothing. \end{cases} \tag{3}$$

The trilateral-based pixel similarity ensures that pixels on either side of the depth discontinuity line are not judged as homogeneous pixels, as shown in Fig. 3(b). Finally, the matching cost of these homogeneous pixels is updated using the riverbed method [7], which can induce the continuous disparity or depth changes in the homogeneous region of an image. Therefore, the riverbed method ensures that most pixels, including those near-depth discontinuity lines, are more inclined to the correct disparities than the incorrect disparities.

C. Depth Discontinuity Lines-Aware SGM

Because SGM demonstrates superior matching precision and computational efficiency [12], the proposed method is an optimization of SGM. The SGM method aims to minimize the global cost function by going along 1-D paths L in eight directions r through the image. Along each path L_r, the minimum cost to reach all disparities of a pixel p on the path is calculated recursively

$$\begin{aligned} L_r(p, d) = C(p, d) + \min(&L_r(p - r, d), L_r(p - r, d - 1) \\ &+ P_1, L_r(p - r, d + 1) + P_1, \\ &\times \min_i L_r(p - r, i) + P_2) - \min_k L_r(p - r, k) \end{aligned} \tag{4}$$

where P_1 and P_2 are the penalty parameters for the current pixel p and the previous pixel in the direction of path r with a disparity difference of one and a disparity difference greater than one, respectively, and $P_2 \geq P_1$. The last subtraction guarantees that $L_r(p, d) < C\max + P_2$. The SGM aggregates the cost paths over all paths on each pixel

$$S(p, d) = \sum_r L_r(p, d). \tag{5}$$

Thus, the disparity for each pixel corresponds to the minimum aggregated cost. Small penalty parameters are used for depth discontinuities to allow for significant changes in the disparities of adjacent pixels. Therefore, an extension penalty tuning algorithm was proposed to tolerate significant disparity changes on both sides of the depth discontinuity lines during the cost aggregation.

After adaptively adjusting the penalty parameters using the gradient, canny, or texture edges (optional), we updated the penalty parameters based on whether neighboring pixels on the cost aggregation path laid on the depth discontinuity line. As shown in Fig. 3(c), if at least one of the current pixel p and the previous pixel $p–r$ in the aggregation direction has a depth discontinuity line, the disparity information passed to the current pixel is no longer reliable, and a smaller penalty parameter $P_2(p)$ should be chosen to allow for disparity mutations. If both p and $p–r$ have no depth discontinuity lines, the penalty parameter $P_2(p)$ was not adjusted. The specific equation is as follows:

$$P_2(p) = \begin{cases} P_2 & p \notin \text{DL and } p - r \notin \text{DL} \\ P_1 & p \in \text{DL or } p - r \in \text{DL}. \end{cases} \quad (6)$$

Finally, disparities around depth discontinuity lines will be more likely to obtain the correct matching result when combined with the trilateral update and penalty tuning.

III. VALIDATION

To validate the effectiveness of the LiDAR and image line-guided stereo matching method (L2GSM), we proposed and evaluated the indoor and aerial stereo images, particularly on artificial structures with many line features. In addition to stereo images, the aerial dataset provides LiDAR data, and the indoor dataset provides reference disparity maps collected by structured lighting scanners. The reference disparity maps of our selected indoor dataset have been proved to be a valid alternative to LiDAR data, owing to its very high accuracy [10].

A. Validation Metrics

We define overall accuracy and edge accuracy to evaluate the disparity maps generated by different DIM methods. The average error and the percentage of outliers [7], calculated by the difference between DIM disparity value and real disparity value, are commonly used overall accuracy indicators. Object edge buffers on the non-object side is an effective method of evaluating edge accuracy [17]. When the absolute difference is greater than 3 pixels between the dense matching disparity d_i and true disparity d_i^m, the pixel i is considered a false match. Mismatched pixels are placed in the set of error matches S_{error}, and other pixels are placed in the set of correct matches S_{correct}. The percentage of mismatched pixels in the edge buffer is defined as follows:

$$\text{badper} = [S_{\text{error}}]/[S_{\text{error}} + S_{\text{correct}}] \quad (7)$$

where $[\cdot]$ represents the number of all elements in the set. We also evaluate the root mean square error (RMSE) of the linear buffer corresponding to the nonobject side. The RMSE is defined as follows:

$$\text{RMSE} = \sqrt{\sum\nolimits_{d_i \in S_{\text{error}} \cup S_{\text{correct}}} \left(d_i - d_i^m\right)^2/[S_{\text{error}} + S_{\text{correct}}]}. \quad (8)$$
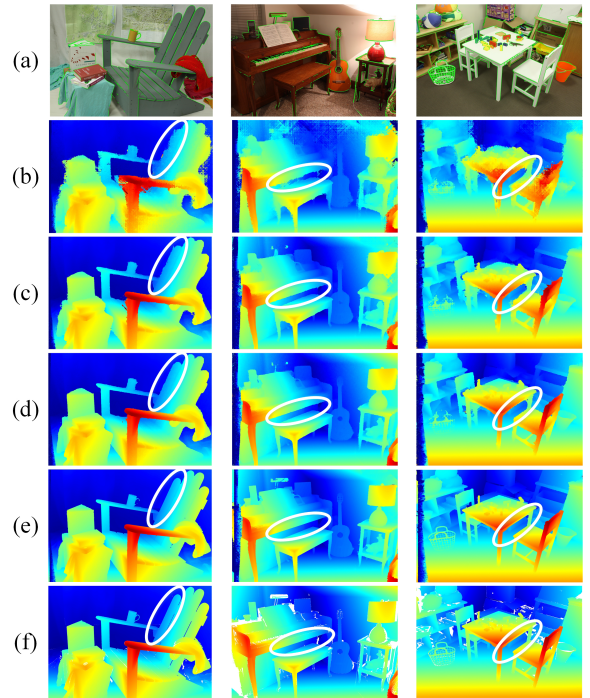

Fig. 4. Experimental results for the Middlebury 2014 dataset. (a) Left rectified image. (b) SGM. (c) Gauss. (d) Diffusion-based. (e) L2GSM. (f) Ground truth disparities.

TABLE I
MATCHING ACCURACY FOR MIDDLEBURY DATASET

Method	Overall accuracy				Edge Accuracy		Time
	>1px	>2px	>3px	Avg/px	badper	RMSE/px	/sec
SGM	30.72%	20.37%	16.26%	3.56	53.63%	0.54	**35.63**
Gauss	9.69%	4.33%	3.30%	1.13	36.85%	0.37	36.85
Diffusion Based	12.07%	2.67%	1.61%	0.88	27.87%	0.28	36.75
L2GSM without cost update	28.44%	17.65%	13.47%	2.51	43.59%	0.44	35.80
L2GSM without penalty tuning	2.95%	1.89%	1.58%	0.59	11.60%	0.12	62.81
L2GSM	**2.57%**	**1.63%**	**1.40%**	**0.50**	**9.29%**	**0.09**	63.37

B. Experimental Results of Indoor Data

We evaluated the L2GSM method first using the Middlebury 2014 dataset [18]. Five percent of the ground truth data were randomly sampled as the LiDAR constraints, and the remaining 95% were used to evaluate the matching accuracy. The sampling percentage is reasonable for the actual data acquisition and is consistent with that of the experiment using the Gauss method [9].

We compared our method with two other state-of-the-art cost volume optimization methods: the Gauss method [9] and the diffusion-based method [10]. A qualitative comparison of disparity maps is presented in Fig. 4. Disparities from small to large are mapped to the pseudo color from blue to red. The original SGM substantially restored the 3-D shape of the object but produced the worst object boundary among these compared algorithms. With the introduction of LiDAR constraints, disparity maps generated by dense matching are improved significantly, as shown in Fig. 4(c)–(e). As shown in the white ellipse area, among all the comparison methods, the object edges of L2GSM are the closest to the object edges of the ground truth disparities.

The statistical results of the overall and edge accuracy, calculated by the remaining LiDAR projection points, are shown in Table I. To perform ablation experiments, we added L2GSM without a cost update and L2GSM without penalty

tuning. The former means that L2GSM only uses depth discontinuity lines-aware SGM. The latter means that L2GSM uses only trilateral updates of the cost volume. L2GSM doubles the computational time compared with the original SGM, but it has less than one-fifth the error rate of the original SGM for all verification metrics. All the LiDAR constraints' SGM methods improved the edge accuracy, and L2GSM is optimal. L2GSM without penalty was better than L2GSM without a cost update in all evaluation metrics in our ablation experiments. We also observed that the matching accuracy of L2GSM without penalty tuning and L2GSM is relatively close. Nevertheless, depth discontinuity lines-aware SGM is essential because L2GSM generates better disparities near the depth discontinuity lines than L2GSM without penalty tuning. Therefore, L2GSM is optimal among all compared algorithms in terms of overall accuracy and edge accuracy.

C. Experimental Results of Aerial Data

We also evaluated the performance of L2GSM by using the Vaihingen dataset, which provides aerial images with a ground sampling distance of 0.08 m and airborne laser scanner data for which the average point density is 4 points/m^2. The images and LiDAR data collected in the area have been accurately registered. The dataset was provided by the German Society for Photogrammetry, Remote Sensing and Geoinformation (http://www.ifp.uni-stuttgart.de/dgpf/DKEP-Allg.html). The matching cost of SGM is the census, and the regularization parameter P_2 of SGM adapts to the intensity gradient.

We selected 15 stereo pairs from 20 images based on geometric configuration and interimage overlap. Due to the significant outdoor light variation, we further compared L2GSM with the light-variation-insensitive pyramid stereo matching network (PSMNet). PSMNet can effectively reduce the matching errors in ill-posed regions by taking advantage of the capacity of global context information [19]. Depth maps of the SGM, Gauss, diffusion-based, PSMNet, and L2GSM methods around building edges are shown in Fig. 5. PSMNet was trained by satellite image datasets. The depth map of the original SGM has many matching errors on the roof, mainly due to matching uncertainty caused by weak textures. The depth map generated by PSMNet does not have many outliers but is not sharp enough at the edges of buildings. With the introduction of LiDAR constraints, the matching results for the low-textured regions are significantly improved. However, compared with the Gauss and diffusion-based methods at building boundaries, our L2GSM method achieves the most accurate matching results. In addition, Fig. 5 shows the elevation profiles of the two white lines on the elevation maps generated by various methods. On the roof (low-textured area), the elevation profiles of the SGM and Gauss methods have false elevation protrusions, and the elevation profiles of the diffusion-based and L2GSM methods vary continuously and consistently. At the building edge, the elevation profiles of the SGM, PSMNet, and diffusion-based methods have different degrees of offset from the building edge. The elevation profiles of the L2GSM method and building edge overlap. Thus, the elevation maps and elevation profiles confirm that the L2GSM method outperforms the SGM, PSMNet, Gauss, and diffusion-based methods in low-textured areas and building edges.

Because the Vaihingen dataset does not have the actual elevation of all pixels, the input of edge accuracy is slightly
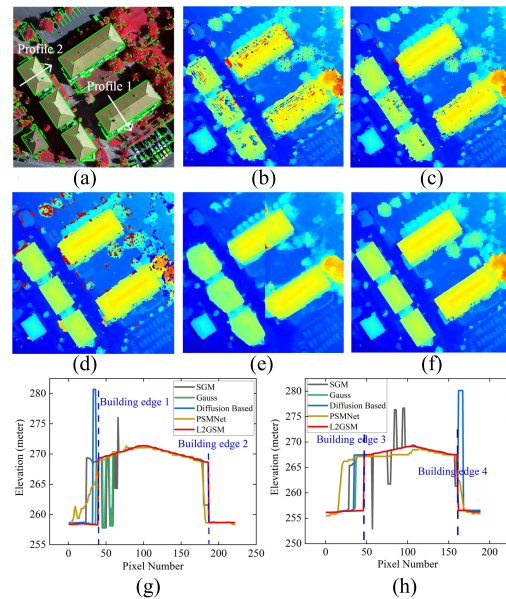

Fig. 5. Experimental results for the aerial stereo image dataset. (a) Depth discontinuity lines. (b) SGM. (c) Gauss. (d) Diffusion-based. (e) PSMNet. (f) L2GSM. (g) Elevation profile 1. (h) Elevation profile 2.
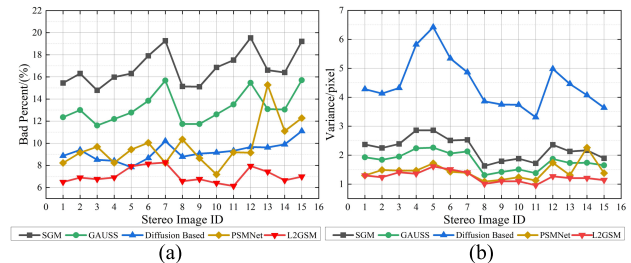

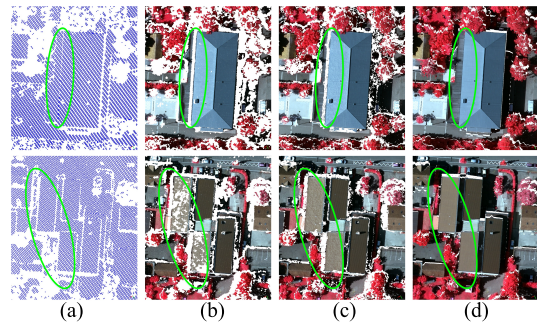Fig. 6. Edge accuracy on multiple aerial stereo images. (a) Percentage of mismatch pixels. (b) Variance.

Fig. 7. LiDAR and DIM point clouds. (a) LiDAR. (b) Metashape. (c) SURE aerial. (d) L2GSM.

adjusted. We calculated the average elevation values on each side of the linear buffer according to the LiDAR projection points. We evaluated the edge accuracy of 15 stereo images. As shown in Fig. 6, the original SGM has the highest number of false matches, and the diffusion-based method has the largest variance among all comparison methods. The accuracy of PSMNet is high on most pairs but poor on a few pairs. Therefore, the bad percentage and variance of L2GSM stabilized at a low level for all 15 stereo pairs, which shows that L2GSM is valuable for preserving the depth discontinuity.

We further evaluated the PSMNet finetuned by aerial stereo pairs and our L2GSM method using the Vaihingen benchmark dataset [20]. About 2% of LiDAR points were randomly sampled from the ground truth data as the input of L2GSM. The

TABLE II
FITTING RESIDUALS AT 3-D LINE FEATURES (UNIT: METER)

Statistics	LiDAR	Metashape	SURE	L2GSM
Average error	0.38	0.43	0.35	**0.32**
Variance	0.15	0.17	0.18	**0.13**

percentage of outliers >1, 2, or 3 pixels of the finetune PSMnet was 31.4%, 20.7%, or 16.1%, while L2GSM, compared with the former, decreased by 3.2%, 1.7%, or 0.6%.

Fig. 7 shows the LiDAR point clouds and the dense matching point clouds generated by our L2GSM method and two state-of-the-art software packages (Metashape and SURE aerial). Metashape and SURE aerial processed stereo images at full resolution and used default values for all matching parameters. To increase the clarity of the visual comparison, we selected two types of building regions. The LiDAR point clouds are sparse and unevenly distributed. From the overall view, Metashape, SURE aerial, and our L2GSM method substantially restored the 3-D structure of the region. Regarding details, Metashape and SURE aerial had some matching errors in low-textured and depth discontinuity regions (e.g., shadows, roofs, or building edges), which were represented as invalid areas of the point cloud. Compared with LiDAR point clouds and other DIM point clouds, the invalid area of the point clouds generated by our L2GSM method is the smallest.

The 3-D straight lines of the buildings, which were measured on the stereo image pair, were used as check features to quantitatively assess the geometric quality of the DIM point clouds [21]. Fitting residuals can relatively reflect the ability of the various methods in describing the nearly ideal lines. Fitting residuals consist of the average error and RMSE obtained from the distance from the dense matching edge points to the adjacent 3-D straight line. Table II shows fitting residuals of many uniformly distributed 3-D line features. Metashape and SURE aerial are close to the fitting residuals of the LiDAR point clouds, which demonstrates the excellent performance of these two packages. L2GSM is better than LiDAR point clouds and dense matching point clouds alone.

IV. CONCLUSION

This letter proposes an L2GSM method that combines high-accuracy LiDAR points and sharp object edges of images to generate accurate and fine-structure point clouds. The experimental results for the indoor and aerial datasets show that our method significantly improves the results of the original SGM and outperforms two state-of-the-art LiDAR constraints' SGM methods. Ablation studies demonstrate the value of a trilateral update of cost volume and depth discontinuity lines-aware SGM in recovering the 3-D structure of low-textured and depth discontinuity regions, respectively. In addition, the 3-D point clouds generated by our L2GSM method outperform LiDAR data and dense matching point clouds generated by Metashape and SURE aerial in terms of completeness and edge accuracy. Therefore, the proposed L2GSM can enhance the 3-D perception capability of LiDAR data and stereo image fusion.

ACKNOWLEDGMENT

REFERENCES

[1] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. J. Comput. Vis.*, vol. 47, nos. 1–3, pp. 7–42, Apr. 2002.

[2] H. Albanwan and R. Qin, "A comparative study on deep-learning methods for dense image matching of multi-angle and multi-date remote sensing stereo-images," *Photogramm. Rec.*, vol. 37, no. 180, pp. 385–409, 2022.

[3] N. Tatar, H. Arefi, and M. Hahn, "High-resolution satellite stereo matching by object-based semiglobal matching and iterative guided edge-preserving filter," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 10, pp. 1841–1845, Oct. 2021.

[4] A. Karami, F. Menna, F. Remondino, and M. Varshosaz, "Exploiting light directionality for image-based 3D reconstruction of non-collaborative surfaces," *Photogramm. Rec.*, vol. 37, no. 177, pp. 111–138, Mar. 2022.

[5] J. Tachella et al., "Real-time 3D reconstruction from single-photon LiDAR data using plug-and-play point cloud denoisers," *Nature Commun.*, vol. 10, no. 1, pp. 1–6, Nov. 2019.

[6] C. Paris and L. Bruzzone, "A three-dimensional model-based approach to the estimation of the tree top height by fusing low-density LiDAR data and very high resolution optical images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 467–480, Jan. 2015.

[7] Y. Zhang, S. Zou, X. Liu, X. Huang, Y. Wan, and Y. Yao, "LiDAR-guided stereo matching with a spatial consistency constraint," *ISPRS J. Photogramm. Remote Sens.*, vol. 183, pp. 164–177, Jan. 2022.

[8] X. Huang, R. Qin, C. Xiao, and X. Lu, "Super resolution of laser range data based on image-guided fusion and dense matching," *ISPRS J. Photogramm. Remote Sens.*, vol. 144, pp. 105–118, Oct. 2018.

[9] M. Poggi, D. Pallotti, F. Tosi, and S. Mattoccia, "Guided stereo matching," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 979–988.

[10] S. S. Shivakumar, K. Mohta, B. Pfrommer, V. Kumar, and C. J. Taylor, "Real time dense depth estimation by fusing stereo with sparse depth measurements," in *Proc. Int. Conf. Robot. Autom. (ICRA)*, May 2019, pp. 6482–6488.

[11] H. Hirschmuller, "Stereo processing by semiglobal matching and mutual information," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 328–341, Feb. 2007.

[12] M. Rothermel, K. Wenzel, D. Fritsch, and N. Haala, "SURE: Photogrammetric surface reconstruction from imagery," in *Proc. LC3D Workshop*, vol. 8, no. 2. Berlin, Germany, 2012, pp. 1–9.

[13] H. Hu, C. Chen, B. Wu, X. Yang, Q. Zhu, and Y. Ding, "Texture-aware dense image matching using ternary census transform," *ISPRS Ann. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 3, no. 3, pp. 1–8, 2016.

[14] Y. Zhang, X. Xiong, and X. Shen, "Automatic registration of urban aerial imagery with airborne LiDAR data," *J. Remote Sens.*, vol. 16, no. 3, pp. 579–595, 2012.

[15] D. Wujanz, M. Burger, M. Mettenleiter, and F. Neitzel, "An intensity-based stochastic model for terrestrial laser scanners," *ISPRS J. Photogramm. Remote Sens.*, vol. 125, pp. 146–155, Mar. 2017.

[16] X. Huang, R. Qin, and M. Chen, "Disparity refinement of building edges using robustly matched straight lines for stereo matching," *ISPRS Ann. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. IV-1, pp. 77–84, Sep. 2018.

[17] R. Qin, M. Chen, X. Huang, and K. Hu, "Disparity refinement in depth discontinuity using robustly matched straight lines for digital surface model generation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 1, pp. 174–185, Jan. 2019.

[18] D. Scharstein et al., "High-resolution stereo datasets with subpixel-accurate ground truth," in *Proc. 36th German Conf. Pattern Recognit. (GCPR)*, vol. 8753, Sep. 2014, pp. 31–42.

[19] J.-R. Chang and Y.-S. Chen, "Pyramid stereo matching network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 5410–5418.

[20] T. Wu, B. Vallet, M. Pierrot-Deseilligny, and E. Rupnik, "A new stereo dense matching benchmark dataset for deep learning," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. XLIII-B2-2021, pp. 405–412, Jun. 2021.

[21] T.-Y. Chuang, H.-W. Ting, and J.-J. Jaw, "Dense stereo matching with edge-constrained penalty tuning," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 664–668, May 2018.