

Super Resolution of Laser Range Data Based on Image-guided Fusion and Dense Matching

Xu Huang^a, Rongjun Qin^{a,b,*}, Changlin Xiao^a, Xiaohu Lu^a

^aDepartment of Civil, Environmental and Geodetic Engineering, The Ohio State University,

^bDepartment of Electrical and Computer Engineering, The Ohio State University, 205 Drees Labs, 2015 Neil Avenue, Columbus, OH, 43210, USA

*Corresponding Author: 470 Hitchcock Hall, 2070 Neil Avenue, Columbus, OH 43210, USA. Tel: +1 614 2924356. Email: qin.324@osu.edu

Abstract

Low-cost/real-time laser range scanner is becoming one of the dominant tools in acquiring accurate 3D point clouds for many smart applications (e.g. automated driving), while the low point density is often the limiting factor for acquiring fine-scale information. On the other hand, stereo/multi-stereo images offer considerably higher-resolution 3D data with low cost, while the image-derived point clouds generally have a higher level of uncertainty. In order to generate accurate, dense point clouds at a low cost, this paper explores a complementary data fusion of the low-resolution-high-accuracy laser range data and the high-resolution (high-res) images, and proposes a super resolution method of laser range data through a novel dense matching framework. In general, we formulate the super resolution as maximizing a posteriori - Markov random field (MAP-MRF) problem in a constrained matching framework, where a two-step strategy is introduced to remove partial inconsistencies between laser range data and images, and the confidence of the high accuracy laser points are propagated through a uniquely designed path in the high-resolution image space, such that a global dense matching algorithm can be externally constrained to yield an accurate, dense and high-fidelity point clouds. We compared the experiment results of the proposed method with the original laser range data and other two super resolution methods of laser range data under aerial, terrestrial, and indoor scenarios. These all demonstrate that the proposed method is capable of producing sub-pixel accuracy, high-fidelity point clouds, even though the density of laser range data is considerably low (hundreds of times lower than the image resolutions).

Keywords

Super Resolution; Laser range data; Image; Fusion; Matching

1. Introduction

Laser range scanner or Light Detection and Ranging (LiDAR), is becoming one of the dominant tools in acquiring 3D information. By actively measuring distance to a surface through time of flight of laser beams in a high-frequency, the laser range scanner is able to collect a vast amount of individually measured high-accuracy point clouds or laser range data. Its characteristics of being robust and accurate have fueled many smart applications including automated mapping, autonomous vehicles, virtual reality, cultural heritage, robotic navigations, etc.

Though promising, the cost and payload for high quality laser range scanner are still relatively high for wide deployment in these applications. For example, the premium laser scanners (e.g. Velodyne HDL-64E, RIEGL VZ-400) being able to generate millions of points per second cost nearly \$80,000 with a payload approximating 12 kg, while such a device limits their applications in civilian mobile platforms, e.g. phone, car, drones. Although there exist low-cost (hundreds to thousands of dollars) and light-weight (dozens to hundreds of grams) laser range scanners such as Canesta EP Devkit, Swiss Ranger SR4500, and RealSense SR300, such sensors are only able to collect sparse point clouds (Yang *et al.* 2007) and are applicable to certain applications such as robotic object avoidance, detection of moving objects and shape completion (Wu *et al.* 2015), while given the resolution limits (average dozens of thousands of points per second), it is not suitable for several high-level vision tasks such as extraction of sub-building level objects, e.g. cars, small architectural features.

Most of the low-cost laser range scanners are equipped with a high-resolution (high-res) camera whose image resolution or pixel density is actually dozens or even hundreds of times higher than laser range scanner. However, the high-res images along with these low-cost laser range scanners were normally used to map textures for generating low-resolution (low-res) models.

A largely ignored fact is that these ready-to-use high-res images are able to form stereo pairs that can be potentially used for generating low-cost, high-density point clouds through dense image matching (DIM) techniques, thus to enhance the resolution of the output for such low-cost system. On the other hand, it is generally known that DIM techniques are scene dependent and the accuracy of the 3D point clouds strongly depends on the resulting image geo-referencing, scene environment, and object types. Hence, point clouds generated from DIM techniques often contain high uncertainties. For instance, image-based point clouds through DIM are normally problematic in regions with high directional reflectivity and poor textures, such as water surfaces, snowfields. Additionally, the state-of-the-art DIM methods are generally not able to handle objects with complex structures, such as fences, wire frames or vegetation brunches. In these scenarios, a laser range scanner performs relatively well, particularly in poor-texture scenario.

Therefore, an optimal solution of generating low-cost, dense, and accurate point clouds requires the complementary data fusion of both data sources for super resolution. This, moreover, yields textured point clouds that can be used for more advanced tasks such as machine learning, object extraction, semantic recognition, and fine-scale reconstruction. This paper particularly addresses this data fusion and super resolution problem by forming it through a novel dense matching framework: we presume the availability of both low-res laser range data and geo-referenced high-resolution images and develop a laser range data super-resolution framework that is able to 1) account for partial inconsistency between both data sources, e.g. temporal changes between the acquisitions, 2) fully utilize the information of laser range data through newly devised non-local paths and most importantly 3) produce high-resolution, high-fidelity texturized point clouds without additional requirement for hardware input. The core algorithm formulates the super resolution as maximum a posteriori - Markov random field (MAP-MRF) problem, where the confidence of high accuracy laser point clouds are propagated through a uniquely designed path in the high-resolution epipolar space of stereo images, such that a global dense matching algorithm can be externally constrained to yield results that preserve both the laser range data fidelity and image data resolution. Our approach is capable of greatly enhancing the resolution of laser range data using higher-resolution images even when both of the data sources are partially inconsistent.

The rest of the paper is organized as follows: Section 2 introduces an overview of related work; Section 3 describes the methodology of the proposed method in detail; Section 4 shows the experimental results; and Section 5 draws the conclusions based on our works.

2. Related Work

In recent years, there have been several attempts to fuse low-res, accurate laser range data and high-res images to produce high-res, high-fidelity point clouds. The basic idea of super resolution is to up-sample the low-res range data by projecting them to a higher-resolution (2D/3D) grid and then to estimate their actual depth or disparity (parallax in the epipolar space) in a 2D scenario or presence in the 3D space. The images can be either used as the guide for extrapolating depth values or to form stereo pairs to generate geometric data for fusion, therefore the laser range data super resolution methods can be categorized into 1) interpolation-based super resolution and 2) stereo-matching-based super resolution.

Interpolation-based super resolution techniques utilize laser points and a single registered image to estimate depths of all the points/pixels in the higher-resolution grids after up-sampling. The basic assumption of these techniques is that homogenous points must share similar depths. The homogeneity between laser points and the pixels is inferred by the intensity/color/reflectance similarities and the distances between them in the image, which is formulated as the weight values of the pixels in the interpolation. Several research works (Andreasson *et al.* 2006, Wang and Ferrie 2015) locally estimate depth for each pixel through a weighted average of depths of surrounding laser points, which is simple and efficient in yielding edge-aware super resolution results, while such algorithms are generally sensitive to noises. Other researches (Diebel and Thrun 2006, Bódis-Szomorú *et al.* 2015) formulated the interpolation-based super resolution as a global Markov Random Field (MRF) problem, normally regarded as more accurate and robust than the local-window-based methods (Andreasson *et al.* 2006). Hosseinyalamdary and Yilmaz (2015) fully utilized the geometry of laser points and the brightness changes of images to recover high-resolution surfaces, respectively, and introduced a cost function to minimize the differences between these surfaces. Their method is capable of preserving boundary and topology of surfaces,

while its result may be suppressed in the presence of pseudo brightness changes (e.g. image texture). In general, these methods are capable of efficiently improving the resolution of the laser point clouds. However, given the nature that the interpolation-based method essentially only utilizes the original laser points as the major information source, its level of improvement is limited when the camera image resolution is an order of magnitude higher than laser point clouds ($> 10\times$) (Yang *et al.* 2007), while this is often the case of many applications such as aerial mapping (normally $25\times$ - $100\times$), indoor reconstruction using low-res laser range scanners ($> 100\times$).

The most relevant work is the matching-based super resolution technique, which uses at least two overlapped images to form a stereo pair and integrates the laser point clouds into an image dense matching framework. The high-fidelity of laser point clouds is propagated to the pixels in the high-res epipolar space of stereo images, such that the matching confidences of these pixels are greatly improved. Different from the interpolation-based methods, the matching-based super resolution is able to take low-res laser point clouds ($>100\times$) as the input and significantly improve their resolution. Several early studies (Bobick and Intille 1999, Yang 2003) on matching-based super resolution were based on the assumption that all the laser points were inliers, and they propagated the confidence of laser points to pixels through local scanning lines. These approaches are simple and efficient in yielding high-res point clouds, while the partial inconsistency between laser points and images were not considered, nor the laser points were fully utilized due to the local propagation paths. To further utilize the reliability of laser points, some researchers used local interpolation (Rengarajan *et al.* 2004, Geiger *et al.* 2010, Wu *et al.* 2011, Liu *et al.* 2015) or global interpolation (Wang *et al.* 2008, Wang and Yang 2011) to build an initial surface from discrete laser points, and then constrained the disparities of pixels to be close to the initial surfaces. This method was able to maintain the accurate geometry from the laser points to a certain extent, while the constraint may negatively impact the fine structures in stereo image matching. In addition, geometric inconsistencies between laser range data and images are an important matter of concern, and some of the existing matching-based super resolution methods are able to accommodate a small number (less than one-tenth) of laser points as inconsistencies by taking them as soft geometric constraints for confidence propagation (Lhuillier and Quan 2002, Kim *et al.* 2005, Geiger *et al.* 2010, Wang and Yang 2011, Huang *et al.* 2015, Liu *et al.* 2015), thus to average out these inconsistencies. However, handling significantly larger inconsistencies between the images and the laser range data (e.g. temporal changes of objects such as building demolishing or rebuilt) remains challenging.

Therefore, our proposed super resolution method for laser range data based on stereo images aims to address the aforementioned issues by 1) fully utilizing the confidence of laser points through a newly devised non-local propagation paths, 2) providing a scheme to remove large amount of inconsistent parts between laser range data and images, and finally 3) producing high-resolution, high-fidelity point clouds to recover 3D information for fine structures.

3. Methodology

In general, our proposed method aims to use the stereo images to produce pixel-wise 3D points under the constraints of the sparsely available laser range data. The stereo images serve as the guidance in the dense matching framework for fusing both datasets. We therefore term our method as super resolution of laser range data based on dense matching framework (SRDM). To incorporate the laser range data into the dense matching framework, the proposed method formulates the super resolution problem as a MAP-MRF problem in a laser range data constrained epipolar stereo matching framework. However, it is expected that inconsistent parts (e.g. temporal changes, occlusions) between laser range data and images may exist, leading to incorrect guidance in the dense matching framework. Hence, to ensure the high accuracy of the final high-resolution disparity image, we introduce a two-step optimization strategy, where the first step optimization is used to eliminate the inconsistent laser points and the second takes the remaining laser points for a more accurate and robust optimization for disparity image.

As an initial step, the images and the laser range data must be rigorously registered prior to super resolution. In this paper, we adopt the multi-feature based adjustment method (Zhang *et al.* 2015) to refine the orientation of the images with the laser range data. This method extracts line and point features in laser range data and images as controls in the adjustment. The refinement is carried by minimizing the projection errors between the projected controls and extracted features,

which is able to achieve sub-pixel accuracy. The refined camera orientation parameters are then used to rectify the images to the epipolar space, reducing the complicated 2D correspondence search problem to 1D in the stereo matching framework. Our proposed super-resolution method consists of two main components: 1) **preprocessing**: project the laser range data onto a higher-resolution epipolar image pairs, which renders sparse laser points with known disparities/depths in an image grid (also called laser range disparity image), the outliers of the laser points caused by flying objects (e.g. birds) or multiple reflections of laser echoes in the laser range data are then eliminated via a statistical filtering; 2) **two-step optimization for high-resolution disparity estimation**: the optimal solution of the super resolution is obtained in two steps. The first step eliminates inconsistent part of the laser range data by comparing it with the estimated disparities of the MAP-MRF matching. The second step utilizes the remaining (consistent) laser range data projected to the epipolar images for MAP-MRF based matching to complete the disparities of all pixels in the higher-resolution image grid (also called dense disparity image) as the super resolution result. A general workflow of the proposed method is shown in Figure 1.

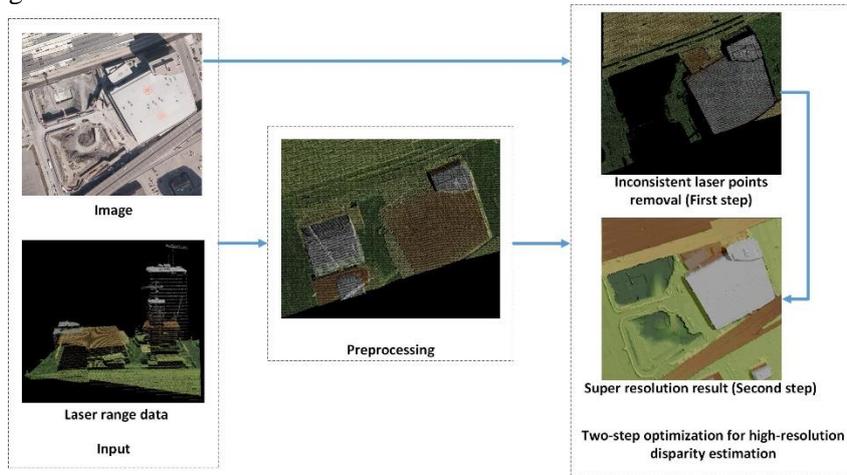


Figure 1. Flow chart of the proposed method.

3.1 Preprocessing

For computational purpose, we project the laser range data onto higher-resolution epipolar stereo images, such that these points are presented as disparities of the stereo epipolar images in Equation (1). Every projected point in the higher-resolution image grids keeps its own disparity while the remaining pixels are invalid, and we call it a laser range disparity image. If several laser points correspond to the same image pixel, the point with the maximal disparity, reflecting the smallest depth, is chosen because others are occluded.

$$d = x_l - x_r \quad (1)$$

where x_l , x_r are column coordinates of corresponding points in the left and right epipolar images, and d is the disparity between correspondences. The laser range disparity image is able to provide predicted disparity search range. However, outliers caused by unexpected flying objects (e.g. birds) or glass roofs often exist in laser range data (Figure 2(a)). We use a histogram based approach (Li *et al.* 2011) to firstly filter out these outliers. Figure 2 shows such an example.

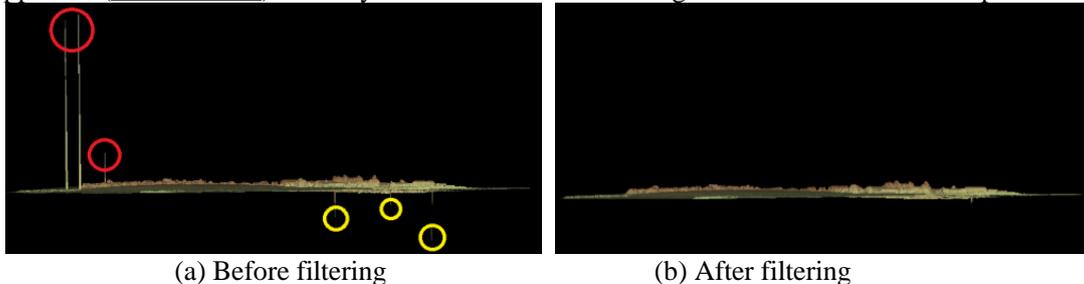


Figure 2. Blunder points filtering in the original laser range data. Red circles and yellow circles in (a) show raised and sunken outliers. (b) shows the laser point clouds after filtering.

3.2 Two-step Optimization for High-Resolution Disparity Estimation

The proposed method formulates the super resolution of laser range data as a MAP-MRF problem in a constrained epipolar stereo matching framework. Since the laser range data will be projected to the epipolar images serving as observations/constraints, it is necessary to eliminate the inconsistencies between the images and laser range data (e.g. temporal changes, occlusions). With the consistent laser range data, we can then propagate their information to optimize higher-resolution disparity image under the MAP-MRF framework. Practically, the inconsistency removal and high-resolution disparity optimization are performed in this MAP-MRF in two steps, with each using slightly different processing strategies: 1) the inconsistency removal uses the laser range data as weak constraints in the MAP-MRF framework for disparity estimation, and then compares the estimated disparities with the original laser range data; 2) the final disparity optimization considers the laser range data as strong constraints in the MAP-MRF framework, to utilize the full potential of the accurate and consistent laser range data. We organize this sub-section by firstly introducing the formulation of super resolution problem (section 3.2.1), and then the solution of the problem through optimization (section 3.2.2). Finally, we describe the implementation of the two-step estimation strategy in the constrained matching framework (section 3.2.3 and section 3.2.4).

3.2.1 Formulation of Super Resolution Problem

Given an laser range disparity image C and a pair of epipolar images $I = \{I_L, I_R\}$ (I_L : left, I_R : right), the goal of the proposed method is to generate a dense, accurate, and high-fidelity disparity image D . Considering that C and I are independent observations, the Bayesian posterior probability over D given I and C is as follows:

$$P(D|I, C) = \frac{P(I|D) \cdot P(C|D) \cdot P(D)}{P(I) \cdot P(C)} \quad (2)$$

where $P(D|I, C)$ is the Bayesian posterior probability; $P(I|D)$ is the conditional probability over I given D measuring pixel-to-pixel the similarities between I_L and I_R ; $P(C|D)$ is the conditional observation of C given D ; $P(D)$, $P(I)$ and $P(C)$ are priors where $P(I)$ and $P(C)$ are dirac delta functions with fixed values since I and C are known, and $P(D)$ is used to model the statistical distributions of D (a priori constraints). Hence, Equation (2) can be rewritten in Equation (3).

$$P(D|I, C) \propto P(I|D) \cdot P(C|D) \cdot P(D) \quad (3)$$

Maximizing the posterior probability of $P(D|I, C)$ in Equation (3) is equivalent to minimize its negative log likelihood, being as a global energy function $E(D)$ (Equation (4)).

$$\begin{aligned} E(D) &= -\ln(P(C|D)) - \ln(P(I|D)) - \ln(P(D)) \\ &= E_{data}(D) + E_{cons}(D) + E_{smooth}(D) \end{aligned} \quad (4)$$

where the data term $E_{data}(D)$ is a function of the disparity measuring the pixel photo-consistency/spectrum difference between the left and right epipolar image under the disparity image D . It calculates, for every pixel in the left epipolar image I_L , its color (or other metric) differences to the corresponding pixel in the right image I_R (navigated by the disparity from D), and then sums these differences across the entire image as the cost to this energy function, as shown in Equation (5). In our method, we employ the combined Histogram of Oriented Gradient (HOG) and Census as the cost metric, given its good capabilities of handling local radiometric differences (from Census) and high matching accuracies in fine textures (from HOG) (Huang *et al.* 2016). In general, the combined HOG and Census metric is a weighted average between HOG and Census metrics in a local small window, as shown in Equation (6).

$$E_{data}(D) = \sum_{p \in I_b/C} \varphi'(p, d_D(p)) \quad (5)$$

I_b is the set of all pixels in a basic image (e.g. the left); I_b/C is the set of pixels in I_b excluding the laser range points in C ; $d_D(p)$ is the estimated disparity of a pixel p ; $\varphi'(p, d_D(p))$ is the matching cost metric of p (e.g. combined HOG and Census in this paper), which is inversely proportional to the photo-consistency probability given $d_D(p)$.

$$\begin{aligned} \varphi'(p, d_D(p)) &= \alpha \cdot \min(\varphi'_{census}(p, d_D(p)), \lambda_{census}) / \lambda_{census} \\ &+ (1 - \alpha) \cdot \min(\varphi'_{HOG}(p, d_D(p)), \lambda_{HOG}) / \lambda_{HOG} \end{aligned} \quad (6)$$

where, φ'_{census} is Census metric relying on the relative ordering of local pixel intensities (Zabih and Woodfill, 1994); φ'_{HOG} is HOG metric relying on an improved HOG features (Huang *et al.*

2016); λ_{Census} and λ_{HOG} are truncated thresholds of φ'_{Census} and φ'_{HOG} , which are empirically defined as 60% of maximums of φ'_{Census} and φ'_{HOG} for limiting the impact of outliers in the cost computation. Given a 5x5 window for cost computation, we set λ_{Census} and λ_{HOG} as 15 and 30 in all experiments. α is a weighting coefficient with range [0, 1], which is defined as 0.5 in all experiments for equally combining φ'_{Census} and φ'_{HOG} . In order to normalize the combined metric, both Census and HOG metrics are divided by λ_{Census} and λ_{HOG} , respectively.

The matching cost in E_{data} is often of high uncertainty, especially in texture-less or repetitive texture regions. Most of the existing matching-based super resolution algorithms used interpolated disparity (from sparse laser range disparity) as constraints to locally improve the robustness of the matching cost for each pixel (Rengarajan *et al.* 2004, Geiger *et al.* 2010, Wang and Yang 2011, Wu *et al.* 2011, Liu *et al.* 2015). However, such constraints for the matching cost often blur fine structures. Hence, we consider a relatively loose constraint on the laser range disparity, which propagate the high confidence of laser range disparity to all pixels using a regularization term.

A regularization term E_{smooth} imposes the smoothness over the entire disparity image with the basic assumption that adjacent pixels with similar color/intensities should share similar disparities, defined as the first order derivatives of \mathbf{D} in a truncated form:

$$E_{smooth}(\mathbf{D}) = \sum_{\mathbf{p}, \mathbf{q} \in \mathcal{N}} P \cdot w(\mathbf{p}, \mathbf{q}) \cdot \min(|d_D(\mathbf{p}) - d_D(\mathbf{q})|, \tau) \quad (7)$$

where \mathbf{p}, \mathbf{q} are neighboring pixels in the high-resolution image grid; \mathcal{N} is the set of adjacent pixels; $w(\mathbf{p}, \mathbf{q})$ is the weight controlling the smoothness constraints based on the intensity differences of neighboring pixels. The weight is inversely proportional to the difference of their intensities, which essentially penalizes the disparity jump between \mathbf{p} and \mathbf{q} when intensities of them are similar. P is a penalty factor controlling the extent of smoothness between adjacent pixels; τ is a truncated threshold that defines the sensitivity of disparity jumps, i.e., smaller τ produces sharper edges while possibly introduce more noises, and vice versa.

The constraint term E_{cons} incorporates the confidences of laser range data in the MAP-MRF optimization. Based on the disparity smoothness constraint, laser range disparity information can be propagated to image pixels to reduce matching ambiguities, which is particularly helpful in texture-less or repeating-texture regions where matching cost metrics are often of high uncertainty. The constraint term calculates, for each laser points its absolute difference (or other related metrics) between the laser range disparity in \mathbf{C} and the estimated one in \mathbf{D} , and then sums these differences of all the laser range points as the constraint term for the global energy function (Equation (8)). When the estimated disparities in \mathbf{D} is highly different from the disparities in \mathbf{C} , these differences become dominant in the global energy function such that the optimal solution has to be constrained to approximate the laser range disparity image for minimizing the global energy function.

$$E_{cons}(\mathbf{D}) = \sum_{\mathbf{p} \in \mathcal{C}} \varphi(\mathbf{p}, d_D(\mathbf{p}), d_c(\mathbf{p})) \quad (8)$$

where \mathbf{p} is a laser point in \mathcal{C} ; $d_c(\mathbf{p})$ is a measured disparity of \mathbf{p} in the laser range disparity image; $d_D(\mathbf{p})$ is the estimated disparity of \mathbf{p} after optimization; $\varphi(\mathbf{p}, d_D(\mathbf{p}), d_c(\mathbf{p}))$ is the constraint cost metric of \mathbf{p} , which measures the disparity difference between $d_D(\mathbf{p})$ and $d_c(\mathbf{p})$. Given that there might be changes between the laser range data and image acquisitions, φ should account for the presence of such inconsistencies by removing inconsistent laser points. Detailed description of φ will be introduced in section 3.2.3 and 3.2.4.

3.2.2 Optimization of the Global Energy Function

The optimization of the global energy function for all image pixels is a typical NP-hard problem, while compromised solutions by breaking the optimization into a collection of sub-optimizations of each pixel in 1D path (e.g. image scanning lines or other related paths) can be performed efficiently in polynomial time. Therefore, several methods utilized the 1D optimization results to approximate to the global energy, such as semi-global matching (SGM) (Hirschmuller 2008) and image-guided non-local matching (INM) (Cigla and Alatan 2013, Pham and Jeon 2013, Sun *et al.* 2014, Cheng *et al.* 2015, Yang 2015). Both methods compute the sub-optimization by using a recursion function to propagate the matching cost between adjacent pixels, while the cost propagation approaches are different. SGM penalizes disparity changes

during cost propagation, using strong penalties for intensity-similar pixels and weak penalties for intensity-different pixels (often corresponding to disparity jumps). However, the weak penalties in disparity jumps may still cause an over-smoothness problem. Different from SGM, INM only keeps disparities same for intensity-similar pixels and does not have weak constraints at all for disparity jumps. Therefore, it can compute accurate disparities in the jumps. However, it may have problems in slanted or curved regions where disparities are not the same. Hence, the entire matching result of INM is not as robust as SGM. In general, the recursion function in cost propagation (also called cost propagation function) of INM is shown in Equation (9) (Cigla and Alatan 2013, Pham and Jeon 2013, Sun *et al.* 2014, Cheng *et al.* 2015, Yang 2015).

$$L_r(\mathbf{p}, d_D(\mathbf{p})) = \varphi'(\mathbf{p}, d_D(\mathbf{p})) + w(\mathbf{p}, \mathbf{p} - r) \cdot L_r(\mathbf{p} - r, d_D(\mathbf{p})) \quad (9)$$

where r is the direction of propagation path; L_r is the propagated cost in direction r ; \mathbf{p} is a pixel in the basic image; $\mathbf{p} - r$ is the previous pixel of \mathbf{p} in direction r . $w(\mathbf{p}, \mathbf{p} - r)$ is inversely proportional to the intensity difference between \mathbf{p} and $\mathbf{p} - r$, which controls the cost propagation between them.

Considering its characteristic of being accurate in disparity jumps, this paper adopts INM optimization and utilizes high confidences of laser range data as constraints to improve the robustness of INM. In general, our method contributes to the INM optimization in two aspects: 1) we improve the cost propagation function of INM by incorporating laser range data as constraints and considering small disparity changes between adjacent pixels (section 3.2.2.1); and 2) we devise a new multi-directional, non-local propagation paths which can fully utilize the confidence of laser range data (section 3.2.2.2).

3.2.2.1 Improved Cost Propagation Function

Matching cost of each pixel often brings ambiguities in the optimization of the global energy function. We utilize laser range data as constraints in the INM optimization and propagate the high confidence of laser range data to image pixels to improve the matching robustness. These constraints are formulated as cost metric measuring the difference between the measured disparity in \mathbf{C} and the estimated disparity in \mathbf{D} , and then we incorporate this cost metric into the cost propagation function of INM to constrain matching, as follows.

$$\begin{aligned} L_r(\mathbf{p}, d_D(\mathbf{p})) &= T[\mathbf{p} \in \mathbf{C}] \cdot \varphi(\mathbf{p}, d_D(\mathbf{p}), d_c(\mathbf{p})) + T[\mathbf{p} \in \mathbf{I}_b/\mathbf{C}] \cdot \varphi'(\mathbf{p}, d_D(\mathbf{p})) \\ &+ w(\mathbf{p}, \mathbf{p} - r) \cdot L_r(\mathbf{p} - r, d_D(\mathbf{p})) \end{aligned} \quad (10)$$

where $T[\bullet]$ is a Boolean function that returns one if its argument is true and zero otherwise; $d_c(\mathbf{p})$ is the measured disparity in \mathbf{C} ; $d_D(\mathbf{p})$ is the estimated disparity in \mathbf{D} ; $\mathbf{p} - r$ is the previous pixel of \mathbf{p} in direction r ; φ' is the photo-consistency cost (e.g. combined HOG and Census in this paper); φ is laser range data based cost. As inconsistencies might exist between laser range data and images, φ should account for the presence of such inconsistencies. Therefore, the formulations of φ are slightly different in inconsistent laser points removal (section 3.2.3) and the high-resolution disparity optimization (section 3.2.4). Details of φ are introduced in these sections respectively.

In addition, the image-guided cost propagation functions (Equation (9) and Equation (10)) only consider matching cost of adjacent pixels at the same disparity level. It does not consider small disparity changes (i.e. ± 1 pixel) between adjacent pixels, while in slanted or curved regions, the disparities of adjacent pixels are the same. In such case, traditional INM optimization may cause mismatches. An example of matching using traditional INM optimization is shown in Figure 3. Figure 3(a) is a typical slanted roof where disparities are continuous with small changes. Figure 3(b-1) is the corresponding matching result using traditional INM optimization (Equation (10)). However, in Figure 3(b-2), the profile map of the matching result in rectangle the red in Figure 3(b-1), shows that the disparities on the roof are forced to be consistent with each other until a large disparity jump, which yields incorrect 3D structures. In order to address such issue in slanted/curved regions, we improve Equation (10) by adding a new constant penalty term at the last for small disparity changes between adjacent pixels, as follows:

$$\begin{aligned} L_r(\mathbf{p}, d_D(\mathbf{p})) &= T[\mathbf{p} \in \mathbf{C}] \cdot \varphi(\mathbf{p}, d_D(\mathbf{p}), d_c(\mathbf{p})) + T[\mathbf{p} \in \mathbf{I}_b/\mathbf{C}] \cdot \varphi'(\mathbf{p}, d_D(\mathbf{p})) \\ &+ w(\mathbf{p}, \mathbf{p} - r) \cdot \min_{k \in [d_p-1, d_p+1]} \{L_r(\mathbf{p} - r, k) + P \cdot T[|d_D(\mathbf{p}) - k| = 1]\} \end{aligned} \quad (11)$$

The last term of Equation (11) returns one if disparities of adjacent pixels change slightly (± 1 pixel) and zero if disparities of adjacent pixels are consistent; P is a predefined penalty factor,

which is used for constraining estimated disparities being smooth. The last term of Equation (11) only lightly penalizes small disparity changes between adjacent pixels, which allows matching in slanted or curved regions.

The matching result using this penalty term (Equation (11)) is shown in Figure 3(c-1). We cropped the red thin rectangle in Figure 3(c-1) to show the disparity trend. The profile map (Figure 3(c-2)) shows that the estimated disparities using the penalty term are continuous smooth with much better matching accuracies in slanted roofs.

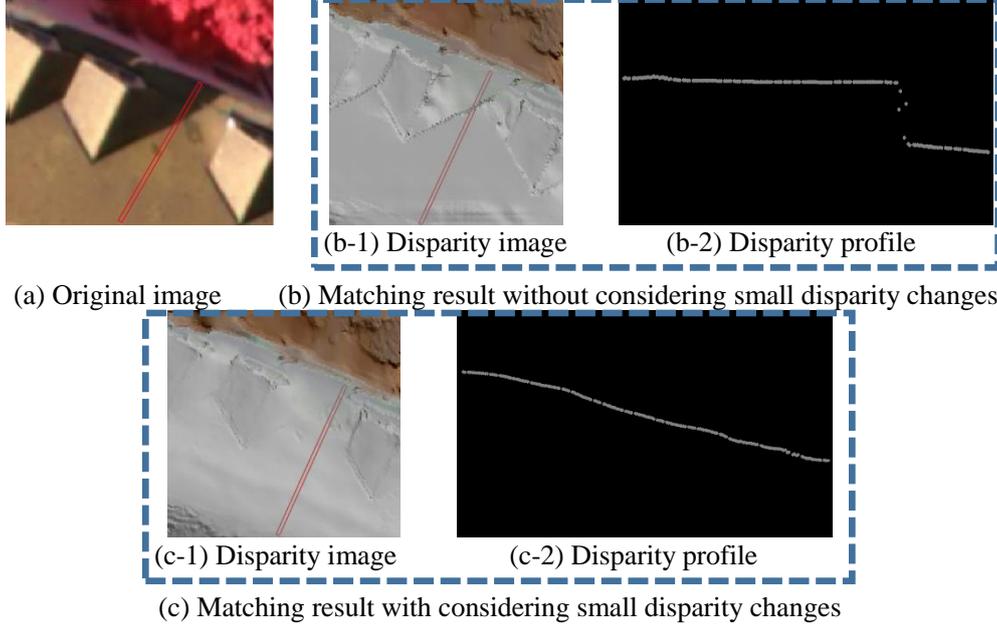


Figure 3. Comparison between the matching results with/without considering small disparity changes in slanted roof. (a) is the original image of a slanted roof; (b-1) is a disparity image without considering small disparity changes; (c-1) is a disparity image with considering small disparity changes. The red narrow rectangles in (b-1) and (c-1) are profiles whose disparity are shown in (b-2) and (c-2), respectively.

According to Equation (11), the matching of image pixels is computed by propagating high confidences (or accurate disparity value) of laser points to them. Given a laser point \mathbf{p} and an image pixel \mathbf{q} , when propagate the confidence of \mathbf{p} to \mathbf{q} , the proportion of the confidence in the matching of \mathbf{q} actually depends on the distance between \mathbf{p} and \mathbf{q} and the intensity difference between adjacent pixels along the propagation path from \mathbf{p} to \mathbf{q} : $\mathbf{p} \rightarrow \mathbf{q}$:

$$W_G(\mathbf{p}, \mathbf{q}) = \prod_{i=\mathbf{p}}^{\mathbf{q}} w(i, i-r) \quad (12)$$

where, $W_G(\mathbf{p}, \mathbf{q})$ measures the proportion of the confidence from \mathbf{p} to \mathbf{q} . Higher $W_G(\mathbf{p}, \mathbf{q})$ means stronger matching constraints from \mathbf{p} to \mathbf{q} . However, due to the resolution limitation of laser points, the distance from a laser point to an image pixel may be dozens to hundreds of pixels, leading considerably small $W_G(\mathbf{p}, \mathbf{q})$. To ensure the effective confidence propagation, the weight w must keep being large (>0.8) along the propagation paths. In most matching approaches, w is computed by a Gaussian kernel: $w = \exp(-\Delta I/\sigma)$ with ΔI : the absolute intensity differences of adjacent pixels, and σ : a smooth factor as its independent variables (often set as 10-20). However, a major issue of the Gaussian function is that it decreases so fast that the value of w is small (<0.4) even when pixel intensities are similar ($\Delta I \approx 10$) in the case of $\sigma = 10$, as shown in Figure 4. The orange dashed line shows the change of Gaussian kernel with increasing ΔI . This issue of the significant declining will reduce the constraint of laser points to other pixels. Therefore, we adopt a quadratic kernel (Equation (13)) whose value is always larger than Gaussian kernel until $\Delta I = 2\sigma$, such as the blue line in Figure 4. Hence, the quadratic kernel can give stronger constraints than Gaussian kernel in intensity homogenous regions.

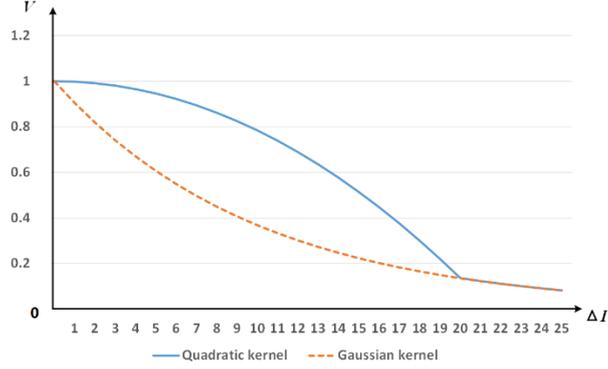


Figure 4. Comparison of the Gaussian kernel and the quadratic kernel in the case of $\sigma = 10$. The horizontal axis shows the intensity difference of adjacent pixels. The vertical axis shows values of the Gaussian kernel and the quadratic kernel.

$$w_q(\mathbf{p}, \mathbf{q}) = \begin{cases} a \cdot \Delta I^2 + 1 & \Delta I \leq 2\sigma \\ Gauss(\Delta I, \sigma) & \Delta I > 2\sigma \end{cases} \quad (13)$$

$$\Delta I = |I_b(\mathbf{p}) - I_b(\mathbf{q})| \quad a = (e^{-2} - 1)/4\sigma^2$$

where w_q is a quadratic weighting function; a is the quadratic coefficient; σ is the smooth factor which controls the sharpness of w_q . We set σ as 10 in all experiments. $Gauss(\bullet)$ is the traditional Gaussian kernel which controls the lower limit of w_q . Equation (13) is a piecewise continuous function. The two pieces $a \cdot \Delta I^2 + 1$ and $Gauss(\Delta I, \sigma)$ have the same value at $\Delta I = 2\sigma$.

3.2.2.2 Non-local Cost Propagation Paths

Good propagation paths can effectively pass the confidence of laser points to the image pixels. The idea of propagating cost along multiple straight-line paths was firstly introduced in semi-global matching (Hirschmuller 2005), where 8 or 16 directions were used. However, the cost propagations along these multiple paths are independent with each other, thus the laser point constraints are not sufficiently utilized, as shown in Figure 5(a). Figure 5(a) shows that there are eight propagation directions through pixel \mathbf{p} , and only laser points along the same path are available to constrain the matching of \mathbf{p} . Although multiple paths can be added up to form a certain global constraint, the laser points beyond paths are not utilized in matching.

Therefore, we propose a new orthogonal and non-local propagation path for fully utilizing all laser points in matching, where the cost propagation of each pixel firstly proceeds along a certain direction r such that only laser points along the direction r are incorporated, such as horizontal direction ($r = 0^\circ$) in Figure 5(b), and then the aggregated cost in r is adopted as the new cost for next aggregation in r_\perp which is perpendicular to r , after which the confidences of laser points can be passed to all other points along paths in r and r_\perp direction. Figure 5(c) shows such an example. The confidence of \mathbf{q} is firstly propagated to \mathbf{s} in the horizontal direction and then propagated to \mathbf{p} in the vertical direction. In addition, its reverse path (i.e. vertical-horizontal direction) and other directions (e.g. $r = 45^\circ$ or 135°) are also computed in this paper.

As compared to SGM, our propagation path is not independent to each other, thus providing a more global propagation. However, not all paths support the cost propagation, as shown in Figure 5. There are in total three paths to pass the cost from \mathbf{q} to \mathbf{p} , namely, $\mathbf{q} \rightarrow \mathbf{s} \rightarrow \mathbf{p}$ follows a first horizontal and then vertical direction; $\mathbf{q} \rightarrow \mathbf{t} \rightarrow \mathbf{p}$ follows a first vertical and then horizontal direction, and $\mathbf{q} \rightarrow \mathbf{p}$ follows the diagonal direction. The path of $\mathbf{q} \rightarrow \mathbf{t} \rightarrow \mathbf{p}$ is unable to support the propagation due to nonhomogeneous pixels in paths (e.g. pixel \mathbf{t}). Both $\mathbf{q} \rightarrow \mathbf{s} \rightarrow \mathbf{p}$ and $\mathbf{q} \rightarrow \mathbf{p}$ can succeed in propagation since all pixels in the path are homogenous. Hence, the matching accuracies of different directions are slightly different. In the following text, we will discuss how these matching results can be best aggregated.

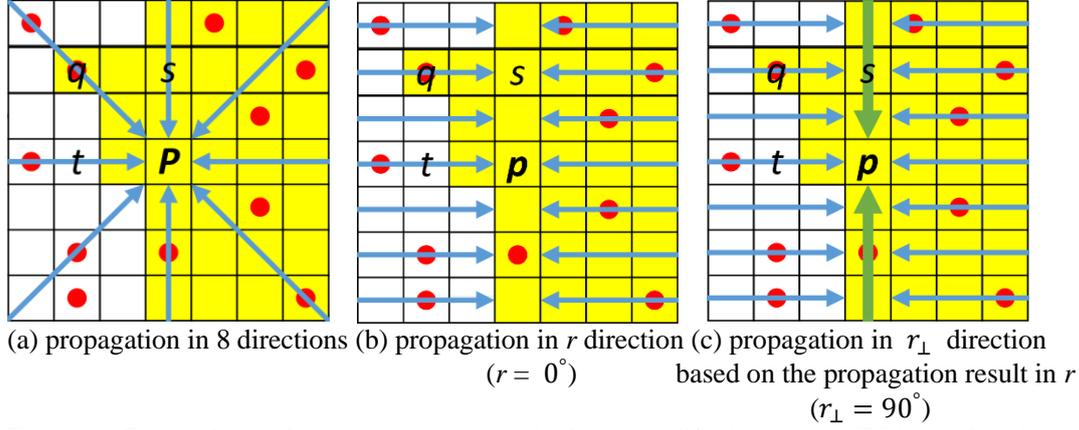


Figure 5. Comparison of cost propagation paths in a simplified example. White and yellow cells represent two different colors in a window. The red points represent laser points. The arrows are the directions of cost propagation. The blue arrows in (b) represent cost propagation in horizontal directions ($r = 0^\circ$), and the green arrows in (c) represent cost propagation in vertical directions ($r_\perp = 90^\circ$) based on the propagation result in horizontal directions.

In general, the cost propagation of each pixel firstly proceeds in both r (here for r we use 0° , 45° , 90° and 135°) and inverse direction of r (termed as $-r$) in Equation (11), and then the propagated cost in these directions are aggregated such that the matching of each pixel can be constrained by laser points in the same direction, as shown in Equation (14).

$$S_r(\mathbf{p}, d_D(\mathbf{p})) = L_r(\mathbf{p}, d_D(\mathbf{p})) + L_{-r}(\mathbf{p}, d_D(\mathbf{p})) - (\mathbb{T}[\mathbf{p} \in \mathbf{C}] \cdot \varphi(\mathbf{p}, d_D(\mathbf{p})) + \mathbb{T}[\mathbf{p} \in \mathbf{I}_b/\mathbf{C}] \cdot \varphi'(\mathbf{p}, d_D(\mathbf{p}))) \quad (14)$$

L_r and L_{-r} are propagated cost in r and $-r$ directions; S_r is the aggregated cost which is the sum of propagated cost in both directions. Both $L_r(\mathbf{p}, d_D(\mathbf{p}))$ and $L_{-r}(\mathbf{p}, d_D(\mathbf{p}))$ contain the cost of \mathbf{p} (φ or φ' , depending on the category of \mathbf{p}). Since the cost of \mathbf{p} appears twice in the sum, the subtracted term in Equation (14) removes this redundancy.

The aggregated cost S_r then is regarded as the new cost for propagation in the perpendicular direction r_\perp , as in Equation (15).

$$L_{r_\perp}(\mathbf{p}, d_D(\mathbf{p})) = S_r(\mathbf{p}, d_D(\mathbf{p})) + w(\mathbf{p}, \mathbf{p} - r) \cdot \min_{k \in [d_p - 1, d_p + 1]} \{L_{r_\perp}(\mathbf{p} - r, k) + 2P \cdot \mathbb{T}[|d_p - k| = 1]\} \quad (15)$$

L_{r_\perp} is the propagated cost in direction r_\perp . It is worth noting that S_r is the aggregated cost of both positive and inverse directions of r . Hence, we double the penalty factor as $2P$ in Equation (15). Then, the cost propagation in the inverse direction of r_\perp (denoted as: $-r_\perp$) can be performed using Equation (15), termed as L_{-r_\perp} .

Finally, the propagated cost in positive and inverse directions of r_\perp are summed to yield the aggregated cost $S_{r_\perp}(\mathbf{p}, d_D(\mathbf{p}))$ in the first r and then r_\perp direction, as shown in Equation (16).

$$S_{r_\perp}(\mathbf{p}, d_D(\mathbf{p})) = L_{r_\perp}(\mathbf{p}, d_D(\mathbf{p})) + L_{-r_\perp}(\mathbf{p}, d_D(\mathbf{p})) - S_r(\mathbf{p}, d_D(\mathbf{p})) \quad (16)$$

The subtracted term in Equation (16) is still needed to remove the redundancy of $S_r(\mathbf{p}, d_D(\mathbf{p}))$ on L_{r_\perp} and L_{-r_\perp} .

Multiple directions including $r = 0^\circ$, $r = 45^\circ$, $r = 90^\circ$ and $r = 135^\circ$ are also incorporated in Equation (17) to allow a more global propagation.

$$S(\mathbf{p}, d_D(\mathbf{p})) = \sum_r S_{r_\perp}(\mathbf{p}, d_D(\mathbf{p})) \quad (17)$$

$S(\mathbf{p}, d_D(\mathbf{p}))$ is the final cost aggregation and we use Winner Takes All (WTA) strategy to get the disparity value, with a quadratic curve fitting (Hirschmuller 2008) to achieve sub-pixel accuracy. Then, a left-right consistency check (LRC) strategy is used to eliminate mismatches and occlusions by comparing disparity images of left images and right images, and the occluded regions can be filled by triangular interpolation (Fowler and Little, 1979).

An example of comparing super resolution results in the newly devised non-local propagation paths in different directions ($r = 0^\circ$, 45° , 90° , 135° , etc.) as well as the traditional 8-direction path is shown in Figure 6. The data in this comparison comes from Piano in Middlebury benchmark (Scharstein *et al.* 2014). The resolution of the original image (Figure 6(a)) is 25 times higher than the laser range data (Figure 6(b)). We use the improved cost propagation function

(Equation (11)) to compute disparity images in traditional 8-direction path (Figure 6(c)) and the newly devised non-local paths (Figure 6(d)-(l)). Among these configurations, the result in the traditional 8-direction path (Figure 6(c)) is worst, due to the insufficient use of laser range data. The results in non-local paths (Figure 6(d)-(l)) are better than that with the 8-direction path, and the matching accuracy of non-local paths can be improved by incorporating more propagation paths.

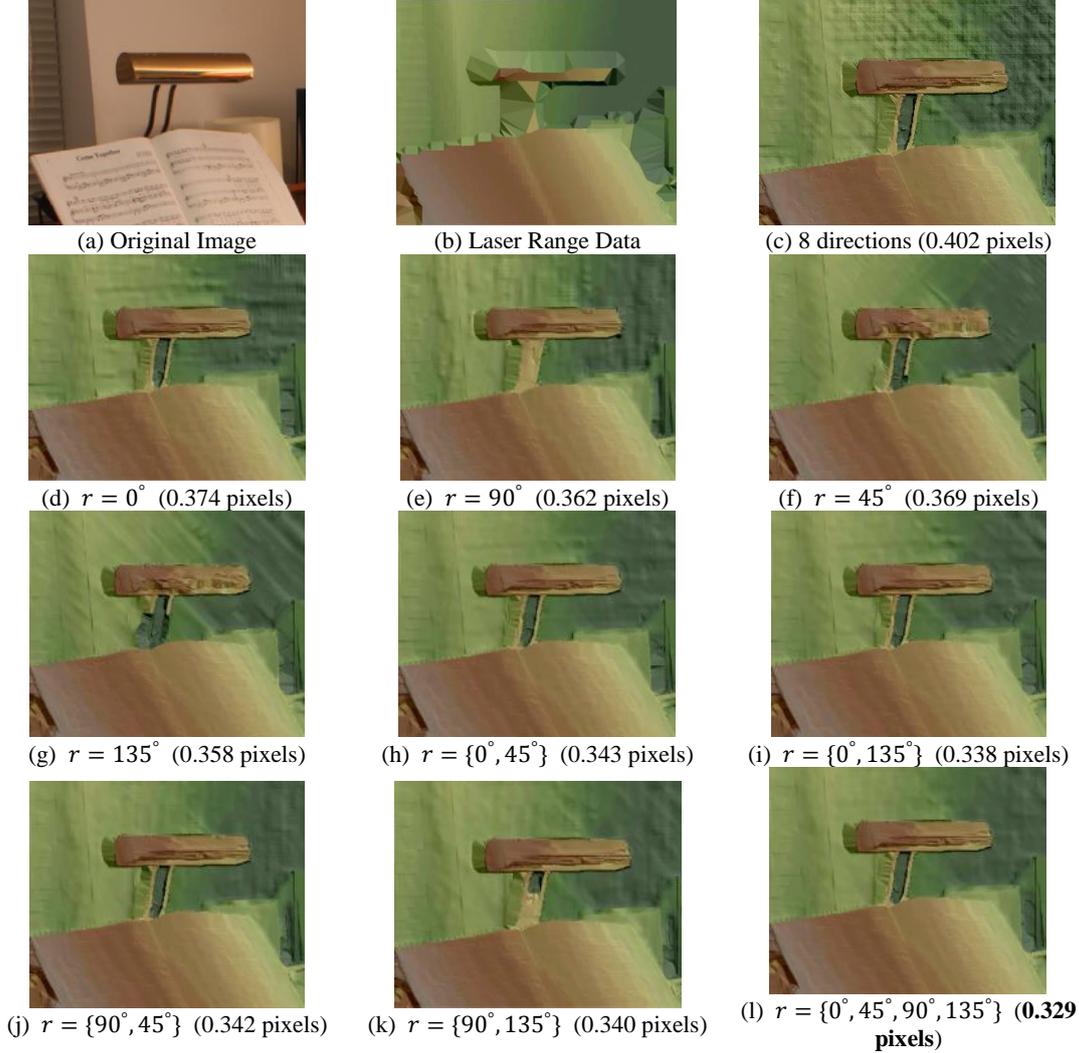


Figure 6. Matching results in different scenarios of propagation paths. The values in round brackets are the corresponding matching accuracies which are computed by averaging differences between disparity images and laser range data.

3.2.3 Inconsistent Laser Points Removal

Inconsistent points between the laser range data and images may exist, e.g. temporal changes, occlusions, and blunders. Therefore, it is necessary to remove these inconsistent points before optimizing the global energy function (Equation (4)). Considering that these inconsistent points often do not satisfy photo-consistency constraints in the matching, we eliminate these points by comparing them with the estimated disparities in stereo matching. We firstly utilize the original laser range data and the images in the optimization of the global energy function, to compute a new disparity image by the proposed optimization method with very weak constraints on the laser data, which may reduce the negative impact of inconsistent laser points. We then compare the laser range data with the new disparity image to eliminate the inconsistent pixels. The remaining laser points are deemed to be reliable, which are used with the images in the optimization again to yield more accurate super resolution results (section 3.2.4).

Firstly, we build a weak constraint term which gives considerably small rewards to the matching cost of points which is far from the disparity of the laser range data:

$$\varphi(\mathbf{p}, d_D(\mathbf{p}), d_c(\mathbf{p})) = \varphi'(\mathbf{p}, d_D(\mathbf{p})) - P_w \cdot e^{-|d_D(\mathbf{p}) - d_c(\mathbf{p})|/\sigma_w} \quad \mathbf{p} \in \mathcal{C} \quad (18)$$

where φ is the laser range constraint; φ' is the matching cost (combined HOG and Census metric); d_c is the measured disparity of a laser point in the laser range disparity image; σ_w is a smooth factor which controls the rewarding range; P_w is a rewarding factor related to the strength of the constrain term. In order to build a weak constraint term, the value P_w cannot be too large. We enforce this weight value being the same as the jump penalty $P_w = P$ in all the experiments in section 4. The rewards are small that the wrong laser point constraints caused by inconsistent points can be corrected by the photo-consistency constraints.

The weak constraint term is incorporated into the global energy function (Equation (4)) which is then solved by the proposed optimization procedure in section 3.2.2 to obtain a new dense disparity image. We then compare the measured disparities of laser points in the laser range disparity image with the generated disparity image, and detect the pixels with inconsistent disparities by thresholding their disparity differences (Equation (19)). The pixels with differences larger than the given threshold are defined as inconsistent pixels thus are eliminated from the laser range data.

$$\mathbf{p} = \begin{cases} inconsistent & |d_D(\mathbf{p}) - d_c(\mathbf{p})| > \delta \\ consistent & |d_D(\mathbf{p}) - d_c(\mathbf{p})| \leq \delta \end{cases} \quad \mathbf{p} \in \mathcal{C} \quad (19)$$

d_D is the estimated disparity in the new disparity image; d_c is the measured disparity in the laser range disparity image; δ is a threshold which measures the disparity consistency. We set $\delta = 2$ in all experiments in section 4.

To speed up the removal process, we apply image pyramids for epipolar stereos. The matching accuracy in top level of image pyramid is usually below ten pixels, while the inconsistency between the laser range data and the estimated disparity image is often larger than dozens or even hundreds of pixels. Hence, the matching result at the top level is sufficient for us to detect inconsistent points. In practice, we build a two-layer pyramid including original level (original stereos) and top level (top layer stereos). To ensure that the resolution of the top layer is still higher than that of laser range data, we define the resolution using the ratio (Equation (20)) between spatial resolution of original images and laser range data.

$$s \leq \lfloor \sqrt{R_I/R_L} \rfloor \quad (20)$$

R_I, R_L are the spatial resolution of original images and original laser range data, respectively; $\lfloor \cdot \rfloor$ is a numerical flooring. With the help of the image pyramids, the running time of the proposed removal strategy can be reduced by s^3 times.

3.2.4 High-Resolution Disparity Optimization

After inconsistent laser point removal, the remaining laser points are deemed reliable with high confidences. We firstly use these laser points to predict disparity search range of each pixel for reducing matching ambiguities, and then construct a strong constraint term to guide matching, thus to compute a high-resolution disparity image. Each pixel is defined as a centric pixel in a predefined square searching window. The disparity search range of them is bounded through laser points with similar intensities to the centric pixel (also called homogenous laser points) in the window. We calculate the maximal and the minimal disparity from these homogenous laser points. The disparity search range of the centric pixel is defined by slightly extending the range between the minimal disparity and the maximal disparity, as shown in the following equation:

$$Range(\mathbf{p}) = [mind - \gamma, maxd + \gamma]$$

$$\begin{aligned} mind &= \min\{d(\mathbf{q}) \mid \mathbf{q} \in Win(\mathbf{p}) \cap |I_b(\mathbf{p}) - I_b(\mathbf{q})| < \varepsilon \cap \mathbf{q} \in \mathcal{C}'\} \\ maxd &= \max\{d(\mathbf{q}) \mid \mathbf{q} \in Win(\mathbf{p}) \cap |I_b(\mathbf{p}) - I_b(\mathbf{q})| < \varepsilon \cap \mathbf{q} \in \mathcal{C}'\} \end{aligned} \quad (21)$$

\mathbf{p} is a centric pixel in the search window; $Range(\mathbf{p})$ represents the predicted disparity search range of \mathbf{p} ; $mind$ and $maxd$ are the maximal and the minimal disparity in the search window; $Win(\mathbf{p})$ is the set of pixels in the search window centered at \mathbf{p} ; \mathcal{C}' is a new laser range disparity image after inconsistent point removal; \mathbf{q} is a laser point in the window; ε is an intensity threshold which decides the intensity similarity between \mathbf{p} and \mathbf{q} ; γ is a value used to extend the disparity range. Only homogenous laser points are involved in disparity range prediction, which helps to maintain high matching accuracies in disparity edges.

The defined search range as in Equation (21) is always much smaller than the overall one of the epipolar stereo. Additionally, it can reduce matching ambiguities and improve the matching efficiency in the MAP-MRF optimization. As the remaining laser points are deemed as reliable,

soft constraints based on the difference of the laser points and the estimated disparities can be used for points where laser range measures are available, whereby the photo-consistency measures are discarded for these points. This retains the laser measures and can only be altered if most of the neighboring pixels with the same color have significantly different disparity. Considering that few inconsistent points may still exist even after inconsistent laser point removal, the strong constraint is truncated by a threshold to achieve an accurate disparity image, as follows:

$$\varphi(\mathbf{p}, d_D(\mathbf{p})) = \min(|d_D(\mathbf{p}) - d_c(\mathbf{p})|, \gamma) \quad \mathbf{p} \in \mathcal{C}' \quad (22)$$

φ is the cost metric of laser points in the constraint term; d is an estimated disparity; d_c is a measured disparity in the laser range disparity image; γ is a threshold to truncate the significant disparity differences.

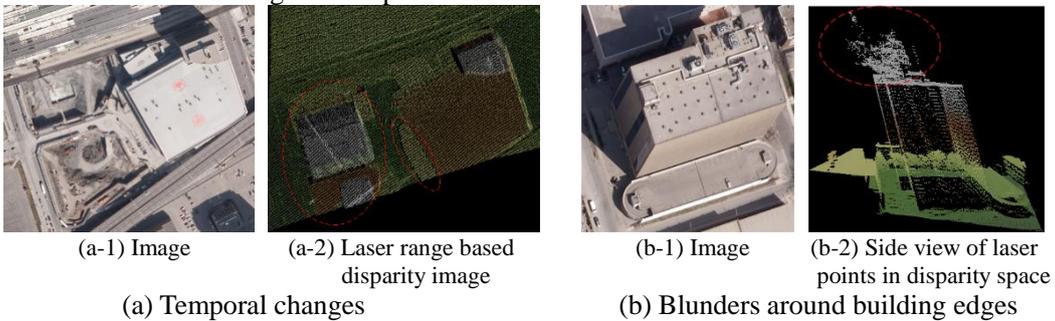
We add this strong constraint into the global energy function and use the proposed optimization method to solve it. The matching of each image pixel is constrained by all laser points with similar intensities through cost propagation along the non-local orthogonal paths. Finally, a dense, accurate disparity image can be produced as the super resolution result.

4. Experiments

We evaluate our proposed method on various laser range datasets including aerial, terrestrial and indoor images. In the initial step, all the image stereo pairs have been rigorously registered to the corresponding laser range data at the sub-pixel level and then rectified to epipolar images before applying the proposed method. The performances of its two major components are assessed: the removal of inconsistent laser points (section 4.1) and the high-resolution disparity optimization (section 4.2). Finally, an example of super resolution in an urban area is used in section 4.3 to compare our proposed method with laser range data and purely stereo image matching.

4.1 Evaluation on Inconsistent Laser Points Removal

This paper used laser range data and aerial images in Toronto provided by the Optech Inc., First Base Solutions Inc., York University and International Society for Photogrammetry and Remote Sensing (ISPRS) WG III/4 (Rottensteiner *et al.* 2012) to test the inconsistent laser point removal step of the proposed method. This dataset was captured by the Microsoft Vexcel's UltraCam-D (UCD) camera (ground sampling distance (GSD): 0.15m) and the Optech airborne laser scanner ALTMORION M (point density: 6 points/m²). The major inconsistent laser points in this dataset include temporal changes (Figure 7(a)), blunders around building edges (Figure 7(b)) and occlusions (Figure 7(c)). Temporal changes were due to the physical change between the capturing time of images and laser range data, e.g. as shown Figure 7(a-2). Blunders around building edges are usually caused by multiple reflections of laser echoes, appearing as a cluster of points with higher disparities, such as the points in the red circle in Figure 7(b-2). Occlusions (Figure 7(c)) are always located in the surfaces which were occluded by higher objects. For example, in Figure 7(c-3), the laser points in region 1 and 2 are invisible in right image (Figure 7(c-2)) and left image (Figure 7(c-1)) respectively, due to the tall building between region 1 and 2. Figure 7(c-4) is a zoomed view of the roof of the tall building. The laser points in region 3 are also blunders around building edges. As far as we know, rare works have addressed such inconsistencies in laser range data super resolution.



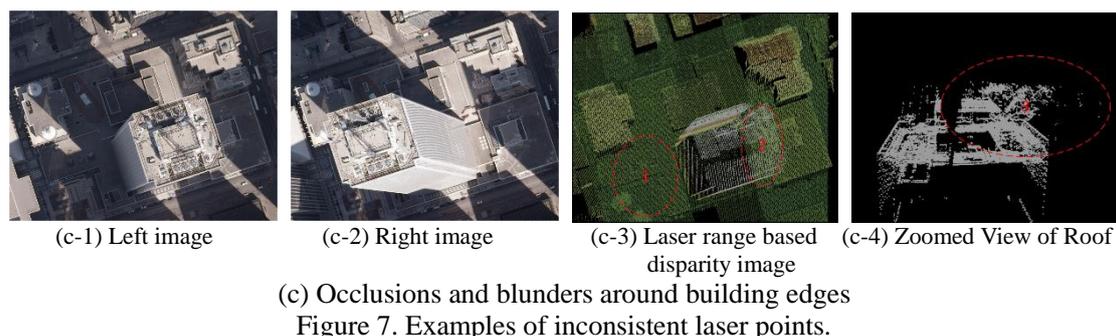
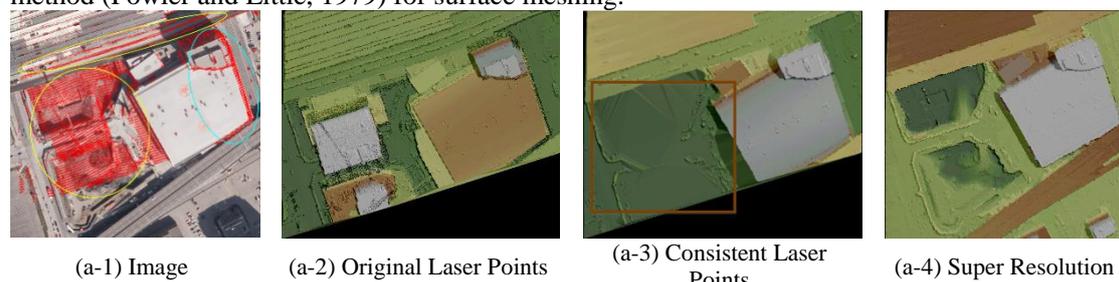
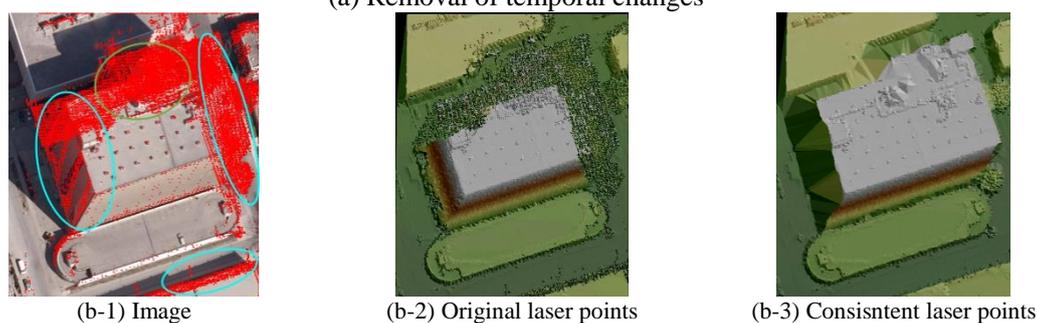


Figure 7. Examples of inconsistent laser points.

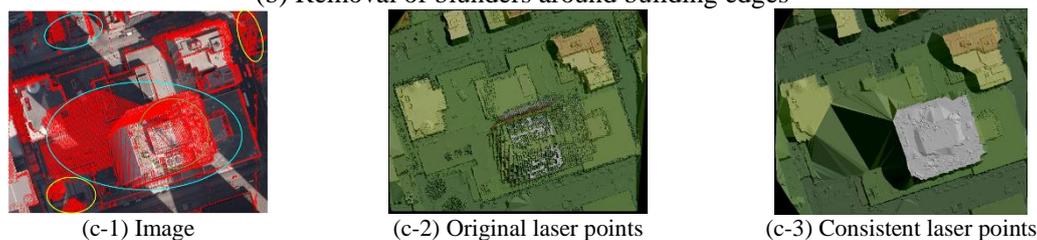
The proposed method is employed to detect and remove the above inconsistent laser points (red points in Figure 8(a-1), (b-1) and (c-1)). Then, the original laser points and the remaining consistent laser points are interpolated through meshing, as shown in Figure 8(a-2), (b-2), (c-2) and (a-3), (b-3), (c-3). In all our experiments, we use the triangulated irregular network (TIN) method (Fowler and Little, 1979) for surface meshing.



(a) Removal of temporal changes



(b) Removal of blunders around building edges



(c) Removal of occlusions

Figure 8. Removal results in different scenarios of inconsistent laser points. E.g. yellow circles: temporal changes; blue circles: occlusions; green circles: blunders around building edges. The brown rectangle in (a-3) represents an invalid region where no laser points exist after the inconsistent point removal.

Figure 8(a-1), (b-1) and (c-1) show that our method is able to detect various types of inconsistent laser points including temporal changes (yellow circles), occlusions (blue circles) and blunders around building edges (green circles). These inconsistent laser points were mixed with consistent ones, leading to ghosting effects in surface meshing (e.g. roofs in (b-2) and (c-2)). After the inconsistent laser point removal, the inconsistent laser points were eliminated and the consistent ones which satisfied the photo-consistency constraints were retained, as shown in Figure 8(a-3), (b-3) and (c-3). Given that the epipolar images naturally defines a 2.5D plane, the ghosting problem can be resolved and the shape of roofs are also recovered (Figure 8(b-3) and (c-3)). The temporal changes in Figure 8(a-2) were removed as well, the missing 3D information

in the brown rectangle and the black regions (not covered by laser range scanner) in Figure 8(a-3) could be recovered by our proposed method in the second step of high-resolution disparity estimation, as shown in Figure 8(a-4).

The removal results are directly related to relative magnitudes between matching cost φ' and the rewarding factor P_w in the weak laser range based constraint (Equation (18)). Too large of a P_w will overweight the constraint of during matching such that the estimated disparities are enforced to be the measured disparities. It is difficult to detect inconsistent laser points in such case, while too small of a P_w will decrease the confidence of laser points resulting in high uncertainties in the matching cost φ' . The matching cost in this paper is combined HOG and Census metric whose range is scaled [0, 1]. Hence, the value of P_w is chosen in this range but not close to 0 or 1. In all the experiments, we adopted $P_w = P = 0.4$ as an empirical value, while the inconsistency removal result is relatively insensitive to the choice of P_w as long as P_w varies from 0.2-0.6.

4.2 Evaluation on High-Resolution Disparity Optimization

As introduced in the related work, the matching-based super resolution methods can be categorized into 1) methods using local propagation paths and 2) methods using initial surface constraints. In this section, we tested the high-resolution disparity optimization step in our proposed method and analyzed the matching accuracies, compared with two typical super resolution methods of the above two categories and three state of the art purely stereo matching methods. The two super resolution methods include the control point constrained global matching (CPGM) (Wang and Yang 2011) and the control point constrained semi-global matching (CPSGM) (Yang 2003). We adopted laser points as control points in both CPGM and CPSGM in the comparison experiment in this section. CPGM uses laser points to build initial surfaces of image pixels, and then formulates the matching as the optimization of a global energy function constrained by these surfaces, and finally computes an optimal solution of the global energy function by graph cuts. CPSGM skips the disparity computation of laser points with the basic assumption that control points/laser points are reliable, and then propagates the high confidence of these laser points to image pixels by dynamic programming in local row directions. For better utilization of laser points, we also added column directions and diagonal directions in CPSGM, namely total 8-direction paths (e.g. Figure 5(a)) were used in CPSGM. We compared our results with three state-of-the-art stereo matching methods: stereo matching using local expansion moves (LocalExp) (Taniai *et al.* 2017), stereo matching using minimum spanning tree for 3D cost aggregation (3DMST) (Li *et al.* 2017) and stereo matching using as-planar-as possible depth map estimation (APAP-Stereo) (Park and Yoon 2006), which rank 2nd, 4th and 5th in Middlebury benchmark (from April 2018). These methods computed reliable matching cost by a trained convolutional neural network (CNN) (Zbontar and LeCun 2016), and then obtain a high-accuracy disparity image through a global optimization strategy.

We used the Piano data (Figure 9(a)) and Moshan relief data (Figure 9(b)) for the comparisons. The Piano data was provided by Middlebury benchmark which were captured using Canon DSLR cameras in medium resolution mode (GSD: 0.001m) and a structure-light system ViewSonic DLP projector with resolution 2-4 times lower than the images (Scharstein *et al.* 2014). The Moshan relief data were captured in Wuhan, China with a Nikon D300 camera (GSD: 0.0009m) and a RIEGL VZ-400 scanner (point density 300,000 points/m²).

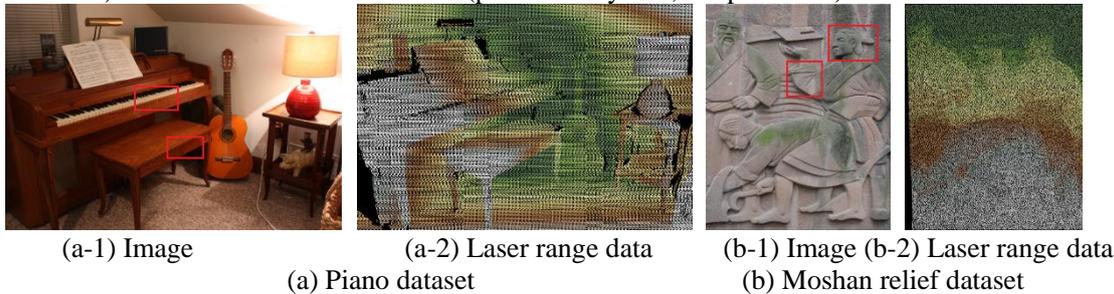
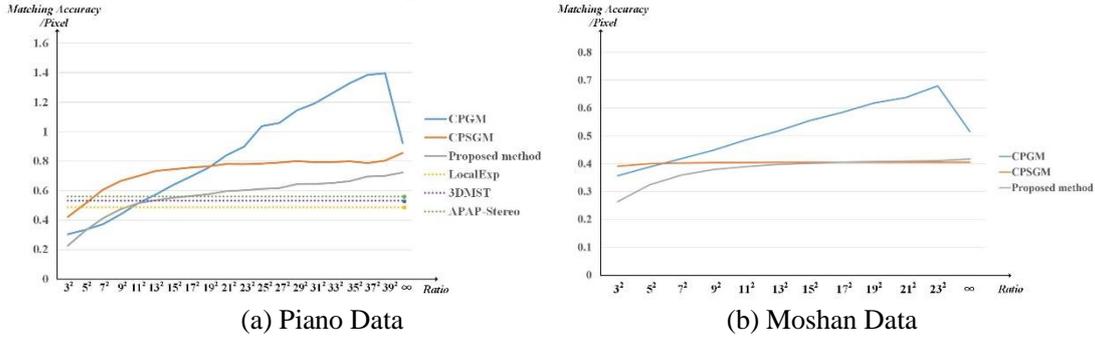


Figure 9. Data used in the matching comparisons.

The original laser point densities in both datasets in Figure 9 are close to the image resolutions. This allows us to analyze the matching accuracies of the three super resolution methods (proposed method, CPGM and CPSGM) by varying the resolution of the laser range data (i.e. down-sampling at different ratios). Low down-sampling ratio (< 10) means a relatively high

1
2

1 density of the laser range data, while high ratio (> 100) means a relatively low density of laser
 2 range data. We down-sample the laser range data (Figure 9(a-2) and (b-2)) with the increasing
 3 ratios: 3×3 , $5 \times 5 \dots \infty$, where ∞ refers to pure image-based matching without the use of laser range
 4 data, and then apply the proposed inconsistency removal strategy on the three super resolution
 5 methods to compute a series of disparity images corresponding to different ratios. We also used
 6 disparity images of state-of-the-art purely stereo matching methods (i.e. LocalExp, 3DMST and
 7 APAP-Stereo) for accuracy comparison on Piano data. The consistent laser points after the
 8 removal in the original resolution are used to evaluate the matching accuracies of all disparity
 9 images by averaging absolute differences between estimated disparities and measured disparities
 10 of these laser points, as shown in Figure 10.

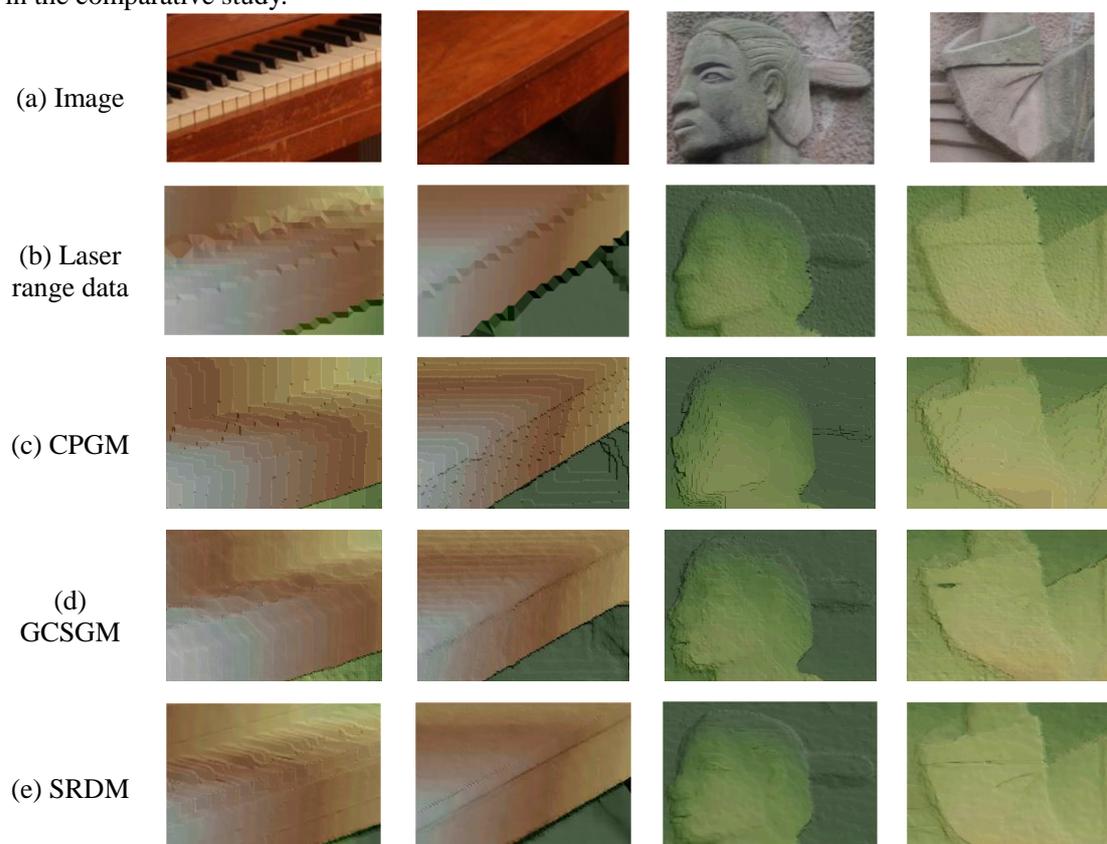


11
 12 (a) Piano Data (b) Moshan Data
 13 Figure 10. Matching accuracies (errors) with the increasing ratio of down-sampling. The
 14 horizontal axis represents the down-sampling ratio of laser range data as compared to the original
 15 image. The vertical axis represents matching accuracies in different ratios. The gray line
 16 represents matching accuracies of our proposed method, the orange line represents matching
 17 accuracies of CPSGM, the blue line represents matching accuracies of GPGM, the yellow dashed
 18 line represents matching accuracies of LocalExp, the purple dashed line represents matching
 19 accuracies of 3DMST, and the green dashed line represents matching accuracies of APAP-Stereo.

20 Figure 10(a) and (b) shows the matching accuracies of the three super resolution methods (i.e.
 21 CPGM, CPSGM and our proposed method SRDM) and the three purely stereo matching methods
 22 (i.e. LocalExp, 3DMST and APAP-Stereo). Ratio = ∞ means that the laser range data is an empty
 23 set ($\mathcal{C} = \emptyset$), meaning a purely image-based stereo matching for CPGM, CPSGM and our proposed
 24 method (denote as SRDM ($\mathcal{C} = \emptyset$)). In this experiment, the matching accuracies of all the three
 25 super resolution methods decreased with increasing down-sampling ratios. CPGM is capable of
 26 computing an accurate disparity image when the ratio is small (below 9×9). Compared with the
 27 purely stereo matching of CPGM at ratio = 3^2 is improved by 67.0% (Piano data) and 30.6% (Moshan data). However, a larger ratio ($> 9 \times 9$) may
 28 cause significant decreases of the matching accuracies, even leading to worse matching accuracies
 29 than the purely stereo matching. The accuracy drop is caused by the initial surface constructed
 30 from discrete laser points in CPGM. When ratio is small (namely, high density of laser range data),
 31 the initial surface can be formulated as robust constraints in CPGM, while when the ratio is large
 32 (namely, low density of laser range data), the initial surface may be unreliable, especially in
 33 curved regions, these surfaces may provide incorrect constraints in CPGM. CPSGM and our
 34 proposed method do not require the initial surfaces. Instead, the matching accuracies of both
 35 methods are improved by passing the high confidence of laser range data to image pixels along
 36 scanning line based paths. The matching accuracies of both methods at ratio $\neq \infty$ are always
 37 higher than their purely stereo matching at ratio = ∞ . However, the matching accuracy
 38 improvement of CPSGM is weak. As compared with its purely stereo matching, the matching
 39 accuracy of CPSGM at ratio = 3^2 is only improved by 50.8% (Piano data) and 3.5% (Moshan data).
 40 The weak improvement is caused by the fact that the 8-direction path used in CPSGM is not able
 41 to fully utilize the laser range data. Our proposed method uses the non-local paths for passing
 42 the high confidence of laser range data. Hence, the matching accuracy improvement is more obvious:
 43 68.7% (Piano data) and 36.7% (Moshan data). Compared with the three state-of-the-art matching
 44 methods (i.e. LocalExp, 3DMST and APAP-Stereo), our proposed method outperforms the others
 45 when ratio $\leq 9^2$. Our proposed method only used simple cost metric and 1D optimization strategy
 46 for matching. However, the comparison shows that given accurate sparse points (ratio $\leq 9^2$), our
 47 proposed method is capable of computing a higher-accuracy disparity image than those
 48

1 state-of-the-art matching methods. It can be seen from Figure 10 (a, b) that a down-sampling ratio
 2 higher than 9^2 does not lead to significant loss of accuracy of our proposed method, thus
 3 preliminarily in this experiment suggesting that if considering higher expense of laser data
 4 corresponds to higher resolution, 9^2 can be a critical scaling (down-sampling) ratio that the
 5 algorithm may best utilize the laser range data with relatively low resolution.

6 The results of these methods at the ratio 5×5 in fine structures (the red rectangles in Figure
 7 9(a-1) and (b-1)) are shown in each row of Figure 11 to compare with laser range data. The
 8 columns in Figure 11 shows matching results of piano keyboard, seat, statue's head and statue's
 9 arm. Figure 11(b) shows the meshed laser range data at ratio 5×5 . All surfaces of the fine
 10 structures (e.g. keyboard and seat) in Figure 11(b) were rather coarse with aliasing boundaries, due
 11 to low laser point density. Figure 11(c) shows the matching results of CPGM. The surfaces in
 12 Figure 11(c) show much higher resolution with clearer boundaries, while the surfaces appear
 13 aliased with integer disparities. The matching accuracies of CPGM at ratio 5×5 in both Piano and
 14 Moshan data were 0.334 pixels and 0.389 pixels (as compared to the original high-resolution laser
 15 range data). Figure 11(d) shows the results of GCSGM which produces smoother surfaces than
 16 CPGM, while the fine structures in Figure 11(d) are still blurry, since CPSGM penalizes disparity
 17 changes between any adjacent pixels, which may lead to over smoothness in fine structures. The
 18 matching accuracies of CPSGM in both Piano and Moshan data were 0.512 pixels and 0.400
 19 pixels. Figure 11(f-h) shows the matching results of the three purely stereo matching methods.
 20 The surfaces in (f-h) were continuous smooth, especially in planar regions (e.g. seat), while the
 21 matching results in fine structures were still not accurate, e.g. results in piano keyboard. The
 22 matching accuracies of LocalExp, 3DMST and APAP-Stereo on Piano data were 0.487pixels,
 23 0.531 pixels and 0.560 pixels. Different from the other methods, our proposed method considers
 24 the fact that fine structures in images often appear as a lot of intensity changes, and it utilizes these
 25 intensity changes to guide the matching. Hence, the surfaces of our proposed method in Figure
 26 11(e) show more details of the objects with higher accuracy, i.e. 0.329 pixels in Piano data and
 27 0.324 pixels in Moshan data. In general, our proposed method can achieve the best disparity image
 28 in the comparative study.



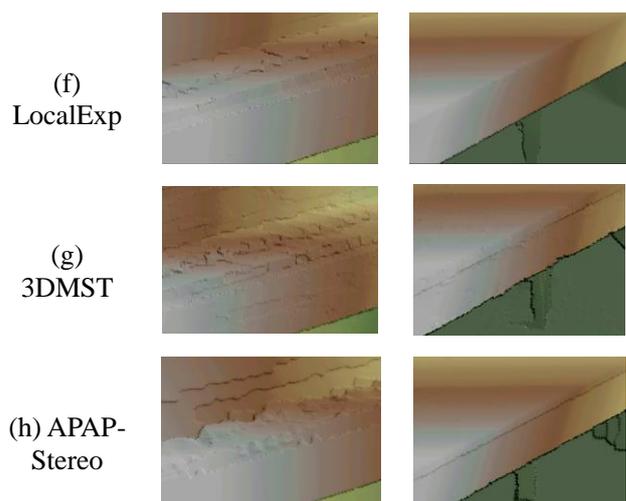


Figure 11. Comparison of indoor or terrestrial fine structures.

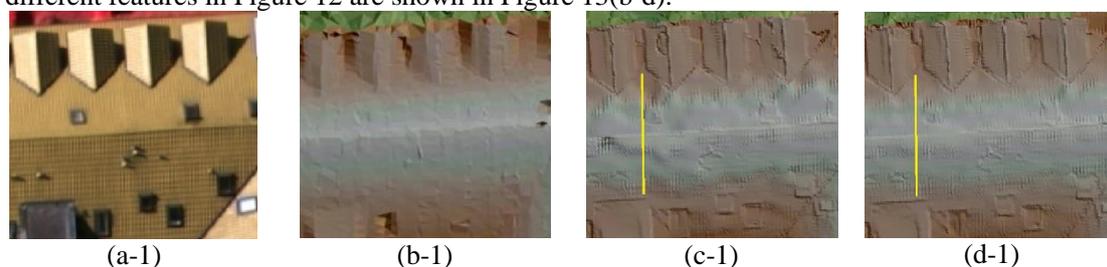
4.3 Example: Super Resolution of Laser Range Data in Urban Area

In this experiment, we apply our method on aerial imagery and laser range data in Vaihingen. The data was provided by the German Society for Photogrammetry, Remote Sensing and Geoinformation (DGPF) (Cramer, 2010), captured by an Intergraph / ZI DMC camera (GSD: 0.08m) and a Leica ALS50 system (point density: 4 points/m²). We collected several fine structures (e.g. shadows, gable roofs, linear features and repetitive textures) for comparison of laser range data, purely stereo matching and the proposed method, as shown in Figure 12.



Figure 12. Typical features in urban area of the aerial dataset, including shadow (blue circled region), gable roofs (yellow circled region), homogenous region (pink circled region) and linear features (green circled region).

The result of the aerial dataset is shown in Figure 13. The original low-resolution laser range data, stereo matching without laser range constraint and the results of the proposed method on the different features in Figure 12 are shown in Figure 13(b-d).



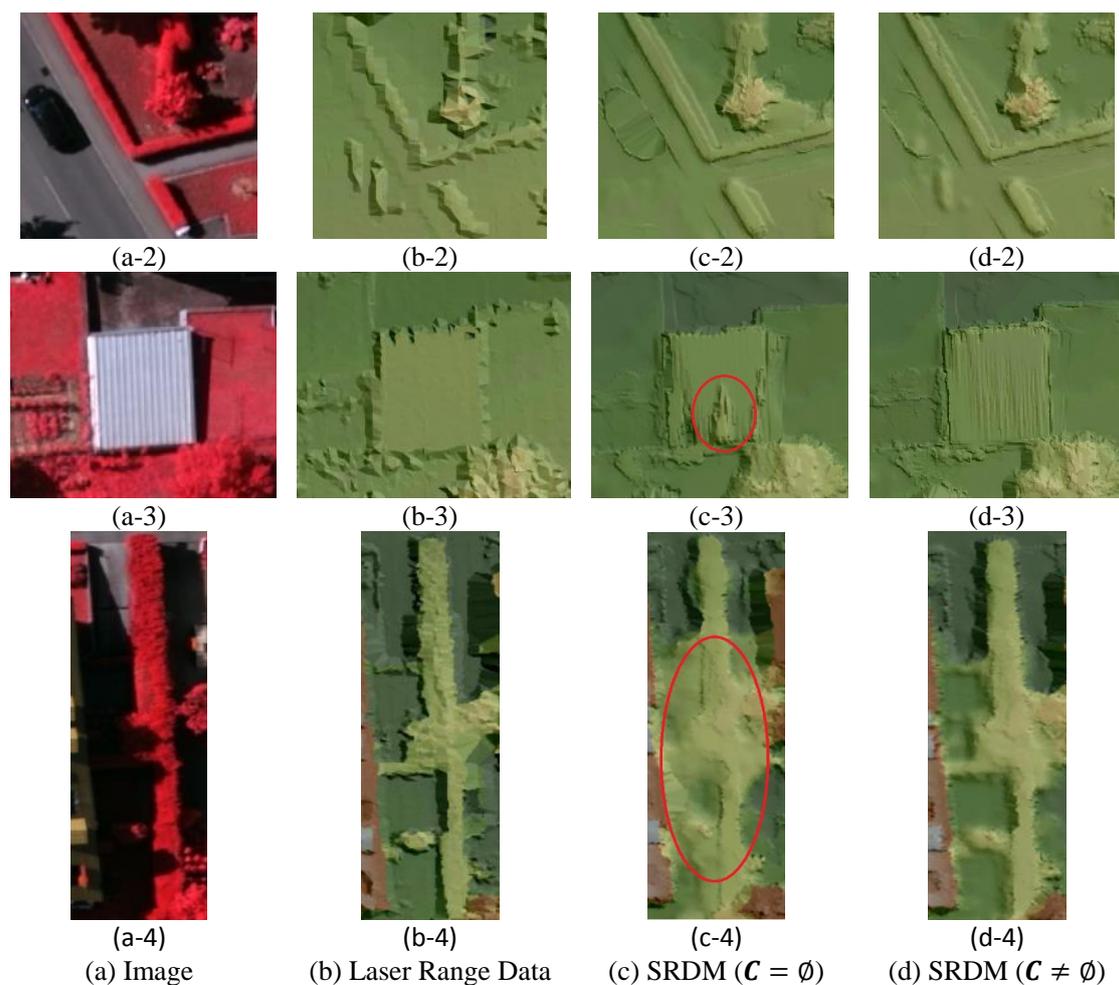


Figure 13. Comparison in fine structures.

1 Figure 13(b) shows the fine structures in laser range data with aliased boundaries and blurred surfaces, due to the low point density. The purely stereo matching method provides point clouds with higher point density (Figure 13(c)), showing details of fine structures. However, there are several mismatches in repetitive textures (red circle in Figure 13(c-3)) and shadows (red circle in Figure 13(c-4)), due to high matching ambiguities in such regions. In addition, matching points on the gable roof in Figure 13(c-1) may yield noises. A disparity profile of the gable roof of Figure 13(c-1) marked by the yellow line is shown in Figure 14. Figure 14(a) shows the disparity profile of laser range data with high accuracies, while the disparity profile in Figure 14(b) shows that the pure matching result is rather noisy in the slant area. This is a typical problem in stereo matching algorithms, which assumes fronto-parallel planes (planes in the scene being parallel to the epipolar line) (Olsson *et al.* 2013), which may introduce bias (up to 1 pixel) in slanted regions. With the laser range data constraint, the proposed method yields a more regularized surface (shown in Figure 14(c)), and statistically more accurate result (0.40 pixels for our method versus 0.75 pixels for purely stereo matching).

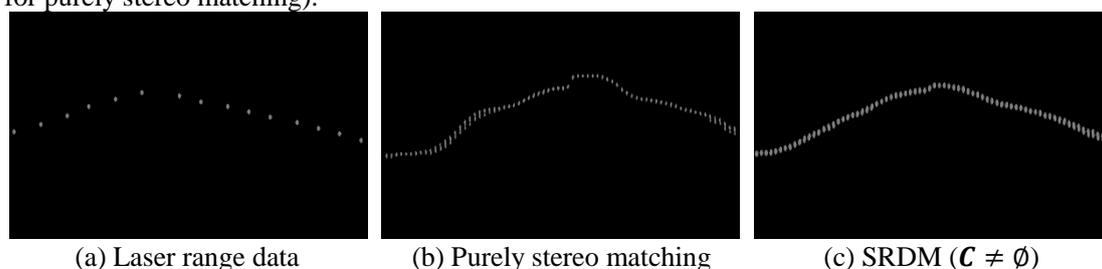


Figure 14. Profile map of the gable roof.

5. Conclusion

1 In this paper, we propose a novel super resolution method of laser range data based on
2 image-guided fusion and dense matching, to improve the resolution of laser range data using
3 high-res stereo images. We formulate the super resolution as the optimization of maximizing a
4 MAP-MRF problem in a constrained matching framework, and introduce a two-step strategy to
5 compute an accurate, high-fidelity super resolution result. The main contributions of our method
6 include: 1) Traditional methods often presume perfect consistency between the laser range data
7 and images, and our method considers the inconsistencies of the two data sources in the
8 optimization framework, thus being able to fuse the data even there exists significant
9 inconsistencies; 2) In the proposed method we devise non-local orthogonal paths for cost
10 propagation so that the reliable information of each laser point can be fully utilized; 3) The
11 proposed method conclusively generates accurate, dense and high-fidelity point clouds at a low
12 cost. Experiments on aerial, terrestrial and indoor datasets demonstrated that our method
13 outperforms the state-of-the-art methods (i.e. CPGM, GPSGM), and in all cases is able to
14 computing disparity images, dense and realistic point clouds at sub-pixel accuracy, even though
15 with the presence of large inconsistencies between laser range data and image data, as well as
16 under cases that the density of laser range data is considerably low (over 100 times lower than
17 image resolutions). As compared to purely image-based stereo matching, our proposed method
18 can efficiently utilize the laser range data to improve the accuracy. For instance, with the aid of
19 laser range data of one-ninth of image resolution, the accuracy is improved at least 36.7%. We also
20 observed that since our method only considers small disparity changes in intensity homogenous
21 regions, mismatches may occur in such region with disparity jumps, thus may incur inaccurate
22 disparity. In our future work, we plan to address such case by considering more robust constraint
23 measure, e.g. using patches through segmentation and constrain segmentation boundaries to obtain
24 more smooth planes on man-made objects.

25
26 **Disclaimer:** Mention of brand names in this paper does not constitute an endorsement by the
27 authors.

28 **Acknowledgement**

29 The work is partially supported by the ONR grant (Award No. N000141712928). We would like to
30 thank Rongyong Huang for providing Moshan relief dataset, the German Society for
31 Photogrammetry, Remote Sensing and Geoinformation (DGPF) for providing the Vaihingen
32 dataset, and the Optech Inc., First Base Solutions Inc., York University, and ISPRS WG III/4 for
33 providing the Downtown Toronto dataset.

34 **Reference**

- 35
36
37 Andreasson, H., Triebel, R., Lilienthal, A., 2006. Vision-based interpolation of 3d laser scans. In:
38 Proceedings of the Proceedings of the 2006 IEEE International Conference on Autonomous
39 Robots and Agents (ICARA 2006), pp. 455-460.
- 40 Bobick, A.F., Intille, S.S., 1999. Large occlusion stereo. International Journal of Computer Vision 33 (3),
41 181-200.
- 42 Bódis-Szomorú, A., Riemenschneider, H., & Van Gool, L., 2015. Superpixel meshes for fast
43 edge-preserving surface reconstruction. In: 28th IEEE Conference on Computer Vision
44 and Pattern Recognition, pp. 2011-2020.
- 45 Cheng, F., Zhang, H., Sun, M., Yuan, D., 2015. Cross-trees, edge and superpixel priors-based cost
46 aggregation for stereo matching. Pattern recognition 48 (7), 2269-2278.
- 47 Cigla, C., Alatan, A.A., 2013. Information permeability for stereo matching. Signal Processing: Image
48 Communication 28 (9), 1072-1088.
- 49 Cramer, 2010. EVALUIERUNG DIGITALER PHOTOGRAMMETRISCHER LUFTBILDKAMERASYSTEME
50 (in German). Retrieved from <http://www.ifp.uni-stuttgart.de/dgpf/DKEP-Allg.html>
- 51 Diebel, J., Thrun, S., 2005. An application of markov random fields to range sensing. In: Proceedings of
52 the Advances in neural information processing systems, pp. 291-298.
- 53 Fowler, R. J., & Little, J. J., 1979. Automatic extraction of Irregular Network digital terrain models.
54 international conference on computer graphics and interactive techniques 13 (2),
55 199-207.

- 1 Geiger, A., Roser, M., Urtasun, R., 2010. Efficient large-scale stereo matching. In: Proceedings of the
2 Asian conference on computer vision, pp. 25-38.
- 3 Hirschmuller, H., 2005. Accurate and efficient stereo processing by semi-global matching and
4 mutual information. In: IEEE Computer Society Conference on Computer Vision and
5 Pattern Recognition, pp. 807-814.
- 6 Hirschmuller, H., 2008. Stereo processing by semiglobal matching and mutual information. IEEE
7 Transactions on pattern analysis and machine intelligence 30 (2), 328-341.
- 8 Huang, X., Yuan, C., Zhang, J., 2015. Graph cuts stereo matching based on patch-match and ground
9 control points constraint. In: Proceedings of the Pacific Rim Conference on Multimedia, pp.
10 14-23.
- 11 Huang, X., Zhang, Y., Yue, Z., 2016. Image-guided non-local dense matching with three-steps
12 optimization. In: ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information
13 Sciences.
- 14 Kim, J.C., Lee, K.M., Choi, B.T., Lee, S.U., 2005. A dense stereo matching using two-pass dynamic
15 programming with generalized ground control points. In: IEEE Computer Society Conference
16 on Computer Vision and Pattern Recognition, pp. 1075-1082.
- 17 Klaus, A., Sormann, M., Karner, K., 2006. Segment-based stereo matching using belief propagation and
18 a self-adapting dissimilarity measure. In: International Conference on Pattern Recognition, pp.
19 15-18.
- 20 Kolmogorov, V., Zabih, R., 2001. Computing visual correspondence with occlusions using graph cuts. In:
21 8th IEEE International Conference on Computer Vision, pp. 508-515.
- 22 Lhuillier, M., Quan, L., 2002. Match propagation for image-based modeling and rendering. IEEE
23 Transactions on Pattern Analysis and Machine Intelligence 24 (8), 1140-1146.
- 24 Li, F., Yu, Z., Wang, B., Dong, Q., 2011. Filtering algorithm for lidar outliers based on histogram and kd
25 tree. In: 4th International Congress on the Image and Signal Processing, pp. 2741-2745.
- 26 Li, L., Yu, X., Zhang, X., Zhao, X., and Zhang, L., 2017. 3D cost aggregation with multiple minimum
27 spanning trees for stereo matching. Applied Optics 56 (12), 3411-3420.
- 28 Liu, J., Li, C., Mei, F., Wang, Z., 2015. 3d entity-based stereo matching with ground control points and
29 joint second-order smoothness prior. The Visual Computer 31 (9), 1253-1269.
- 30 Mozerov, M.G., Van De Weijer, J., 2015. Accurate stereo matching by two-step energy minimization.
31 IEEE Transactions on Image Processing 24 (3), 1153-1163.
- 32 Olsson, C., Ulen, J., Boykov, Y., 2013. In defense of 3d-label stereo. In: Proceedings of the computer
33 vision and pattern recognition, pp. 1730-1737.
- 34 Papadakis, N., Caselles, V., 2010. Multi-label depth estimation for graph cuts stereo problems. Journal
35 of Mathematical Imaging and Vision 38 (1), 70-82.
- 36 Pham, C.C., Jeon, J.W., 2013. Domain transformation-based efficient cost aggregation for local stereo
37 matching. IEEE Transactions on Circuits and Systems for Video Technology 23 (7), 1119-1130.
- 38 Park, M.G. and Yoon, K.J., 2016. As-planar-as-possible depth map estimation. IEEE Transactions
39 on Pattern Analysis and Machine Intelligence. (submitted)
- 40 Rengarajan, R., Yoon, J., Shan, J., 2004. Triangulation based hierarchical image matching for mars dem
41 generation using moc na stereo images. In: Proceedings of the Proceedings of the Annual
42 Convention of the ASPRS, 12pp.(on CD-ROM).
- 43 Rottensteiner, F., Sohn, G., Jung, J., Gerke, M., Baillard, C., Benitez, S., & Breitkopf, U., 2012. The ISPRS
44 benchmark on urban object classification and 3D building reconstruction. In: ISPRS Annals of
45 the Photogrammetry, Remote Sensing and Spatial Information Sciences, pp. 293-298.
- 46 Scharstein, D., Hirschmüller, H., Kitajima, Y., Krathwohl, G., Nešić, N., Wang, X., Westling, P., 2014.
47 High-resolution stereo datasets with subpixel-accurate ground truth. In: Proceedings of the
48 German Conference on Pattern Recognition, pp. 31-42.
- 49 Scharstein, D. and Szeliski, R., 2014. Middlebury stereo vision page,
50 <http://vision.middlebury.edu/stereo/>. (Accessed 03 December, 2017)
- 51 Sun, X., Mei, X., Jiao, S., Zhou, M., Liu, Z., Wang, H., 2014. Real-time local stereo via edge-aware
52 disparity propagation. Pattern Recognition Letters 49, 201-206.
- 53 Tanai, T., Matsushita, Y., Naemura, T., 2014. Graph cut based continuous stereo matching using locally
54 shared labels. In: IEEE Conference on Computer Vision and Pattern Recognition, pp.
55 1613-1620.

- 1 Taniai, T., Matsushita, Y., Sato, Y., & Naemura, T., 2017. Continuous 3D Label Stereo Matching
2 using Local Expansion Moves. *IEEE Transactions on Pattern Analysis and Machine*
3 *Intelligence*.
- 4 Wang, L., Jin, H., Yang, R., 2008. Search space reduction for mrf stereo. In: *European Conference on*
5 *Computer Vision*, pp. 576-588.
- 6 Wang, L., Yang, R., 2011. Global stereo matching leveraged by sparse ground control points. In: *IEEE*
7 *Conference on the Computer Vision and Pattern Recognition*, pp. 3033-3040.
- 8 Wang, R., Ferrie, F.P., 2015. Upsampling method for sparse light detection and ranging using
9 coregistered panoramic images. *Journal of Applied Remote Sensing* 9 (1), 095075-095075.
- 10 Wu, B., Zhang, Y., Zhu, Q., 2011. A triangulation-based hierarchical image matching method for
11 wide-baseline images. *Photogrammetric Engineering & Remote Sensing* 77 (7), 695-708.
- 12 Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X., & Xiao, J., 2015. 3d shapenets: A deep
13 representation for volumetric shapes. In: *Proceedings of the IEEE conference on*
14 *computer vision and pattern recognition*, pp. 1912-1920.
- 15 Yang, Q., 2015. Stereo matching using tree filtering. *IEEE Transactions on Pattern Analysis and*
16 *Machine Intelligence* 37 (4), 834-846.
- 17 Yang, Q., Wang, L., Yang, R., Stewénius, H., Nistér, D., 2009. Stereo matching with color-weighted
18 correlation, hierarchical belief propagation, and occlusion handling. *IEEE Transactions on*
19 *Pattern Analysis and Machine Intelligence* 31 (3), 492-504.
- 20 Yang, Q., Yang, R., Davis, J., Nistér, D., 2007. Spatial-depth super resolution for range images. In:
21 *Proceedings of IEEE Conference on the Computer Vision and Pattern Recognition*, pp. 1-8.
- 22 Yang, Y.-H., 2003. Fast stereo matching using reliability-based dynamic programming and consistency
23 constraints. In: *Ninth IEEE International Conference on Proceedings of the Computer Vision*,
24 pp. 610-617.
- 25 Zabih, R., & Woodfill, J., 1994. Non-parametric local transforms for computing visual
26 correspondence. In: *European conference on computer vision*, pp. 151-158.
- 27 Zhang, Y., Xiong, X., Zheng, M., Huang, X., 2015. Lidar strip adjustment using multifeatures matched
28 with aerial images. *IEEE Transactions on Geoscience and Remote Sensing* 53 (2), 976-987.
- 29 Žbontar, J., & Lecun, Y., 2016. Stereo matching by training a convolutional neural network to
30 compare image patches. *Journal of Machine Learning Research* 17 (65), 1-32.
- 31