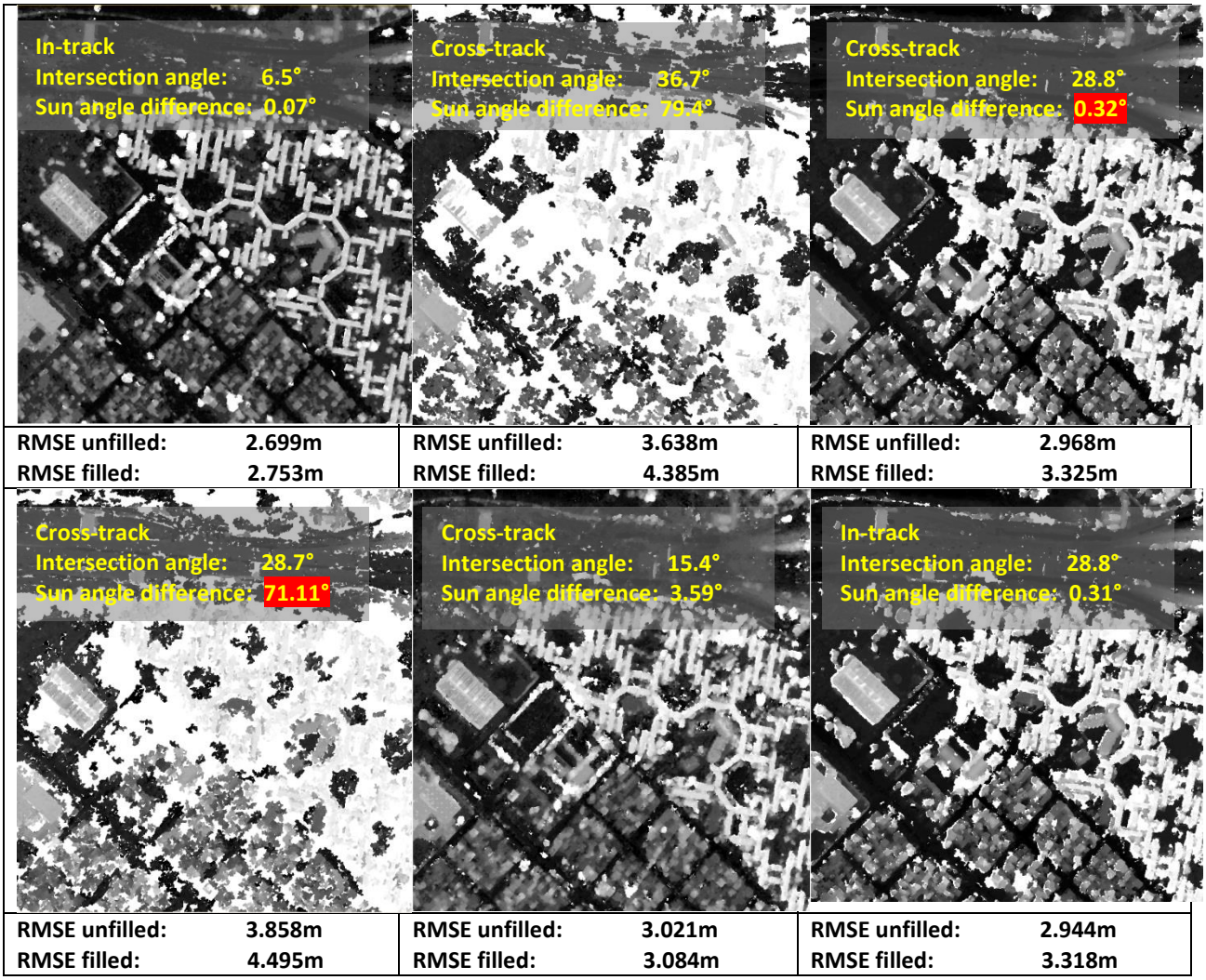
A Critical Analysis of Satellite Stereo Pairs for Digital Surface Model Generation and A Matching Quality Prediction Model

Rongjun Qin[a, b]

[a] Department of Civil, Environmental and Geodetic Engineering, The Ohio State University, 218 Bolz Hall, 2036 Neil Avenue, Columbus, OH, 43210, USA.

[b] Department of Electrical and Computer Engineering, The Ohio State University, 205 Dreese Labs, 2015 Neil Avenue, Columbus, OH 43210, USA.

Graphical Abstract:



Digital surface models generated using Semi-global matching (SGM, with Census as the cost metric) from different stereo pairs with a sample of different meta-information. We demonstrate that the meta-information for stereo pair is critical. In addition to well-known factors such as intersection angles, the sun angle plays a significant role in dense matching results (as highlighted in red). It is shown in the last images of the two rows that the stereo pair being “in-track” or “cross-track” does not significantly impact the results much as long as their meta-parameters are similar.

Abstract: The geometric analysis and data acquisition of satellite photogrammetric images are often regarded as a direct extension of traditional aerial photogrammetry, with the only difference being the sensor model (linear array vs. central perspective). The intersection angle (or base-height ratio) between two images is seen as the most important metadata of stereo pairs, which directly relates to the base-high ratio and texture distortion in the parallax direction, thus both affecting the horizontal and vertical accuracy. State-of-the-art DIM algorithms were reported to work best for narrow baseline stereos (small intersection angle), e.g. Semi-Global Matching empirically takes 15-25 degrees as "good" intersection angles. However, our experiments found that the intersection angle is not the only determining factor, as the same DIM algorithm applied to stereo pairs of the same area with similar and good intersection angle may produce point clouds with dramatically different accuracy (demonstrated in the graphical abstract). ***This raises a very practical and often asked question: what factors constitute a good satellite stereo pair for DIM algorithms?*** In this paper, we provide a comprehensive analysis on this matter by performing stereo matching using the very typical and widely-used Semi-Global Matching (SGM) with a Census cost over 1,000 satellite stereo pairs of the same region with different meta-parameters including their intersection, off-nadir, sun elevation & azimuth angles, completeness and time differences, thus to offer a thorough answer to this question. ***Our conclusion has specifically outlined an important yet often ignored factor – the Sun-angle difference to be one decisive in determining good stereo pair.*** Based on the analytical results, we propose a simple idea by training a support vector machine model for predicting potential stereo matching quality (i.e. potential level of accuracy and completeness given a stereo pair). Experiments have shown that the model is well-suited and generalized for multi-stereo 3D reconstruction, evidenced by a comparative analysis against three other strategies: 1) pair selection based on an example patch where partial ground-truth data is available for computing a priori ranking 2) based on intersection angles and 3) based on a recent algorithm using intersection angle, off-nadir angle and time intervals. This work will potentially provide a valuable reference to researchers working on multi-view satellite image reconstruction, as well as for practitioners minimizing costs for high-quality large-scale mapping. The trained model is made available to the academic community upon request.

Keywords: 3D Reconstruction, Dense Image Matching, Digital Surface Model, Satellite Photogrammetry, Rational Polynomial Functions, Matching Quality Indicator.

1. Introduction

The increasing availability of very-high-resolution (VHR, meter/sub-meter level) spaceborne imaging sensors has driven great interest in acquiring useful geospatial dataset at a global scale. These sensors running 24/7 collect a vast amount of data, generating multiple views of places of interest over the earth surface. Rapid and fully automated 3D reconstruction from these data is extremely useful for geospatial information extraction and global situational awareness. Although recovering 3D information such as digital surface models (DSM) from such data are not new in the geo-community, its practical potential was only brought up in recent years, thanks to the advanced development of dense image matching (DIM) algorithms, allowing for automatically producing LiDAR (Light Detection and Ranging) comparable dense point clouds from images (Gehrke et al., 2010). Due to its relatively low cost and the potential advantage in repeated acquisitions, such 3D contents offer a huge potential for large-scale and wide-area monitoring (Qin et al., 2016), object modeling and land cover studies. However in practice, it is often questioned that the results of the reconstruction are not ideal and unpredictable, even with the most advanced methods appearing in the leaderboard of the benchmark testing data in the computer vision community (Geiger et al., 2012; Scharstein and Szeliski, 2014). These benchmark-testing datasets were captured under careful camera network design and lighting configurations that have minimal impact on the dense matching results (Scharstein and Szeliski, 2002), while such well-designed data

acquisition scenario are often not available in practical cases. For example, most of the satellite images in archives are incidental collections where factors for good stereos are often not considered.

Normally the geometric processing of the satellite imagery is regarded as a direct extension of aerial perspective images, with the only differences being the geometric/camera model. We argue that there are dramatic differences between satellite and aerial photogrammetric image acquisition in terms of their characteristics that impact the DIM: In typical aerial photogrammetric acquisition missions, e.g. using Unmanned Aerial Vehicle (UAV), or aerial platforms, the acquisition can be controlled to capture ideal image blocks with minimal lighting differences and good geometric configurations. The temporal lighting difference is usually not a factor of concern as the acquisitions are done in a comparatively short period. Here good geometric configurations refer to designed acquisition patterns with good overlaps leading to the optimal base-high ratio for narrow-baseline stereo algorithms. In the case of satellite mapping, the acquisition is much more restrictive in terms of: 1) camera networks; 2) acquisition time interval; 3) atmospheric effects; the flying path will need to follow the orbits, and given the fact that most of the satellite platforms carry linear array cameras (near parallel projection in the flying direction), the stereoscopic overlap is realized through steering the camera orientations in the orbit. Normally the time interval of overlapping images in the same track (the same orbit pass: *in-track*) is in a matter of minutes (Dowman and Michalis, 2003), while for overlapping images off the track (in a different orbit pass: *cross-track or incidental collection*) it is in a matter of days, months or even years. Therefore, both the geometric and time constraints bring difficulties in digital surface model (DSM) generation: On one hand, it requires accurate steering of the camera to yield expected convergent images; on the other hand, the long acquisition interval comes along with possibilities of physical changes on the ground, and scene being illuminated under different lighting conditions, and additionally the atmosphere reflection/refraction causes the image digital numbers more sensitive to viewing angles. Therefore, if these factors are not considered, it is not surprising that the accuracy performance of stereo reconstruction algorithms reported in academic studies on a particular dataset can be of little referencing value to their performances on other (new) datasets.

Many of existing studies correlate the dense matching results with the intersection angles (thus base-high ratios) primarily for in-track satellite stereo pairs (Carl et al., 2013; d'Angelo et al., 2014; Zhu et al., 2008), and this is in line with traditional aerial photogrammetry principals. However, when considering satellite stereos aforementioned characteristics (especially for cross-track), good intersection angles as the factor for stereo pair selecting might not be necessarily sufficient for DSM generation. An example is shown in ***Figure 1:*** it can be seen that the left and middle datasets have similar intersection angles but result in dramatically different DSM in quality (RMSE (Root-mean-squared-error) of 4.5m vs. 3.32m). The middle and right datasets have different intersection angles while both of the DSM reveals visually much better results as compared to the first.

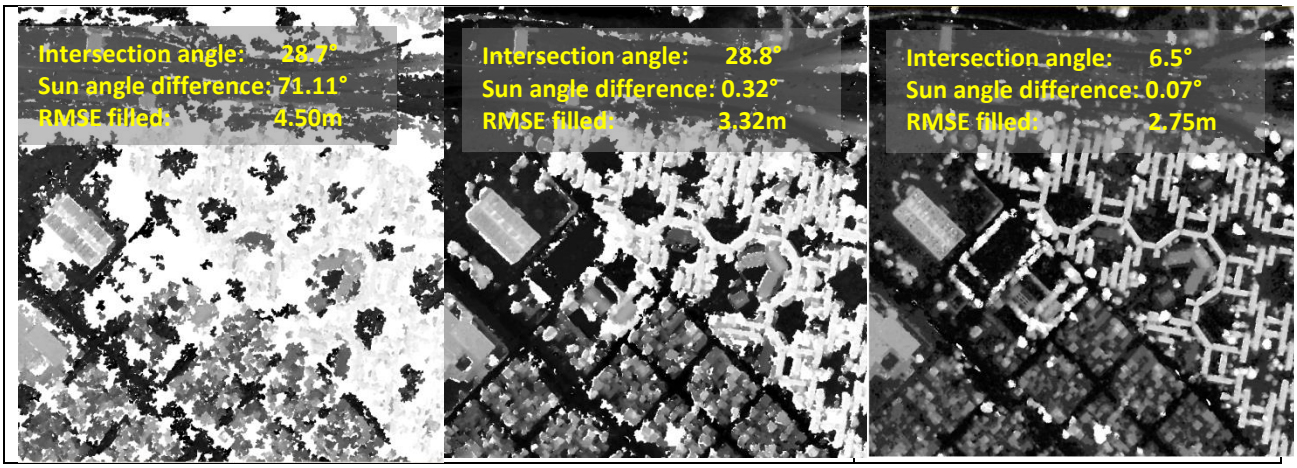


Figure 1. DSM with a GSD (ground sampling distance) of 0.3 meters generated from three pairs with different and similar intersection angles (however with different sun angles); RMSE filled refers to the Root-mean-squared error of the interpolated DSM.

In a work by (Facciolo et al., 2017), the authors investigated a few more parameters including the off-nadir angles and acquisition time intervals on a multi-view stereo satellite benchmark dataset (Bosch et al., 2016a), and concluded that maximal off-nadir angle and the acquisition time interval are also strongly correlated with the resulting accuracy. However, it was shown in a previous work (Zhu et al., 2008) that time intervals might not be critical, as long as the stereo geometry of cross-track data are satisfactory. This is also concluded by a recent work (Krauß et al., 2018) that it is feasible to use cross-track images for DSM reconstruction even with long acquisition interval (more than a year). In their work, they evaluated approximately 136 generated DSM generated from 14 images against ground-truth LiDAR data, and have reported that both intersection angles and the difference in sun angles play important roles in the resulting DSM. However, with simple statistical analyses, they only suggested qualitative and simple pair selection criterions based on educated estimates of angle ranges, and independent tests were not yet performed on other datasets to verify the effectiveness of these criterions. A recent coding challenging focusing on multi-view stereo (MVS) 3D reconstruction from satellite images, has concluded in the winners' workshop that how to select the stereo pairs is the key, which was done manually through trial and error (IARPA, 2016; Qin, 2017). Existing works performing satellite stereo reconstruction normally assume as is and the focuses were more on actual DIM processing. Available resources for performing analysis on the quality of satellite pairs is limited, and the current literature as discussed above are not ready to provide quantifiable conclusions on correlating the quality of stereo pairs with their meta-parameters.

To achieve a comprehensive analysis, this paper investigates critical parameters/metadata characterizing the geometric and radiometric/lighting properties of the stereo data, to demonstrate the association between these parameters and the stereo reconstruction results. We take advantage of a recently available public dataset that contains 50 satellite images and LiDAR point clouds over the same region (details are in ***Section 2***). With a permutation of 1,200 possibilities (significantly more samples than those in prior works), we are able to run stereo reconstruction and compare them with the ground-truth LiDAR data for over 1,000 satellite stereo pairs, the metadata of which are diverse enough to draw conclusive association of critical stereo parameters, including off-nadir angles, intersection angles, sun elevation azimuth angles, etc. and their combinations with the final DSM accuracy. Based on the analyzed results, we build support vector machine models to predict the level of quality of a satellite stereo pair that may yield highly accurate and complete DSM. This can be of critical importance for MVS satellite 3D reconstruction and provides important measures on how to select good satellite stereo pairs

in a large volume of satellite image archives. The rest of the paper is organized as follows: ***Section 2*** provides a brief introduction of the satellite dataset used for the meta-parameter analysis; ***Section 3*** introduces our analysis methodology and experiment setup, and ***Section 4*** presents the experiment and analysis results. Based on the analysis in ***Section 4***, ***Section 5*** introduces our trained support vector machine model using a set of meta-parameters. ***Section 6*** concludes this by summarizing our general findings and advising future works.

2. Dataset

The satellite images used in this work are the multi-view benchmark dataset from John's Hopkins University Applied Physics Lab's (JHUAPL) (Bosch et al., 2016b; Bosch et al., 2017), containing worldview 3 images over this area across two years with a total of approximately 50 images (***Figure.1***(a)), with the area coverage of roughly 150 km^2. They are taken under various conditions containing in-track and cross-track stereos with the ground resolution around 0.3-0.5 meters, with 8-band multispectral information available. Each of these 50 images is provided with their associated parameters including the acquisition off-nadir and azimuth angle, sun elevation & azimuth angle, acquisition time etc. All these datasets were performed a level-2 correction (Ortho-ready, images are projected to a mean elevation plane) such that the image distortions are minimal.

In theory, any two images may form a stereo pair for 3D reconstruction. Of course, there are possibilities that some of them have singular geometric configurations, for example, pairs with extremely small intersection angles (e.g. < 3 degrees). In total this dataset (with 50 images) yields 1,200 possible stereo pairs with a wide range of geometric configurations. This benchmark dataset has provided a means for validation: over one-third of the region (i.e. 50 km^2), the highly accurate airborne LiDAR data are available for ground-truthing. Thus, a root-mean-square error (RMSE) can be computed based on the stereo-reconstructed DSM and LiDAR data, such that the meta-parameters of the stereo pairs (i.e. intersection angles, off-nadir angles, sun elevation & azimuth angles, etc., see a full list in ***Table 1***) can be associated with the resulting accuracy of the DSM for analysis.

Among these 1,200 stereo pairs, there exist a small number of invalid stereo pairs, e.g., very small intersection angle, dramatic season changes in the scene. After eliminating them, we ended up with

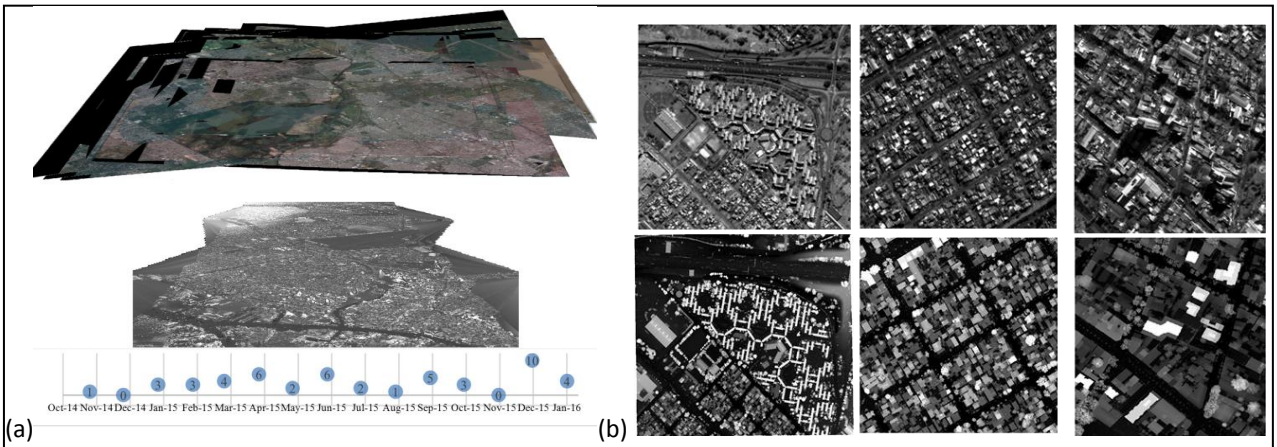


Figure 2. **(a)** A visualization snapshot of part of the 50 images with panchromatic band (multispectral bands are available) across two years, and the acquired LiDAR point clouds. The sample frequency is summarized in the lower part of the figure. **(b)** The selected regions of interest. First row: panchromatic images, second row: DSM generated from the LiDAR point clouds. The first column (a): Test region 1, moderate height commercial building region; the second column (b): Test region 2, low residential buildings; the third column (c): Test region 3, high-rise commercial buildings.

approximately 1,180 stereo pairs. The RSP (RPC stereo processor) (Qin, 2016b) software is used to generate the DSM. RSP implements a hierarchical semi-global matching method (Hirschmüller, 2008) using the Census cost metric, which presents the state-of-the-art operational development of dense image matching techniques. To analyze the performances of the DSM generation under different scenarios, we selected three regions of interest (1 X 1 km^2) to perform the analysis, including scenes containing 1) moderately high commercial buildings; 2) low residential buildings and 3) high-rise commercial buildings. These regions were associated with the cropped DSM generated from LiDAR point clouds as the ground-truth for evaluation. A snapshot of these three regions is shown in ***Figure 2***(b).

Each stereo pair comes along with meta-information, and we evaluate the meta-information of a stereo pair by assessing its resulting DSM accuracy, where we compare both the unfilled (non-interpolated) and filled DSM against the ground-truth LiDAR DSM. A selection of relevant meta-parameters and accuracy measure is listed in Table 1, where we explain each of the parameters the rationale of incorporating them for this study. It should be noted that these parameters mainly consist of recorded or derived angles of the satellite platform or the sun illumination directions.

Table 1. A list of original and derived meta-parameters and accuracy measure for the experiment

Symbol	Description	Rationale
θ_{in}	The average intersection angle, computed from the satellite metadata	A traditional factor of concern in terms of triangulating accuracy
θ_{max_n}	The maximal off-nadir angle of the two images	We expect that the higher off-nadir angle, the more façade content being in the image, the higher chance the matching to be affected by occlusions and shadows
$\theta_{s\_e\_min}$	The smaller sun elevation angle when the two images were taken	The smaller the sun elevation angle, the larger occlusions it might lead
$\theta_{s\_e\_dif}$	The difference of sun elevation angle of the two images	Different sun angles illuminate different texture patterns to the ground. A large difference in sun angles might affect the matching results.
$\theta_{s\_a\_dif}$	The difference of sun azimuth angle of the two images	The same as above
θ_{s_dif}	$\sqrt{\theta_{s\_a\_dif}{}^2 + \theta_{s\_e\_dif}{}^2}$	We assume the difference of both elevation and azimuth angles might serve as a single indicator on how the sun angle difference impacts the matching results. Note this is not a physical angle, rather an indicator incorporating sun illumination different for both elevation and azimuth directions. This is naturally correlated with $\theta_{s\_e\_dif}$ and $\theta_{s\_a\_dif}$
$mon_{\_dif}$	Number of months between two acquisition	The interval between two acquisitions might refer to significant physical change of the ground, or seasonal changes.
$M_{cpltness}$	The percentage of matched pixels (after left-right consistency check)	This is an intermediate level and post-DIM parameter; We expect that normally higher completeness in matching may imply a better generated DSM. The unmatched regions are normally those eliminated by left-right consistency check (Hirschmüller, 2005).
Accuracy Measure		
$A_{unfilled}$	RMSE value of the non-interpolated DSM	The RMSE computation follows a rigorous registration to the ground-truth DSM, and this evaluates the accuracy of the originally derived measurements
A_{filled}	RMSE value of the filled DSM	The unmatched areas are filled by using triangulation-based interpolation. The RMSE computation follows a rigorous registration to the ground-truth DSM. This evaluates takes into account the overall quality of the DSM with an equal amount of points.

3. Methodology and Experiment Set-up

We perform linear regression analyses to interpret the association between the meta-parameters and the accuracy of the data. For the three test regions (***Figure 2***), we first take the 1,180 pairs of data for each of the region and compute their DSM using a fully automated approach as introduced in (Qin, 2017), with the core algorithm being Semi-Global Matching (SGM) with Census as its cost metric (Zabih and Woodfill, 1994). Each stereo pair is independently processed through a "relative orientation + dense matching" procedure, and for each DSM, we co-register them using a simplified surface matching

(Gruen and Akca, 2005) to eliminate the systematic errors in orientation for RMSE calculation. Every stereo outputs two DSMs, the unfilled DSM that retains the 3D measurement in the grid, and a filled DSM using (Delaunay) triangle-based interpolation (for details, see (Qin, 2016a)). The unfilled DSM presents the evaluation the accuracy of the originally matched 3D points, while the filled DSM presents the evaluation the accuracy of the final product. The filled DSM inherently incorporates the completeness of the matching, since the interpolated areas tend to be less accurate and will affect the final accuracy if the DSM contains large unmatched regions. The two accuracy values, $A_{unfilled}$ and A_{filled}, are correlated by each of the meta-parameters listed in ***Table 1*** (θ_{in}, θ_{max_n}, $\theta_{s\_e\_min}$, $\theta_{s\_e\_dif}$, $\theta_{s\_a\_dif}$, θ_{s_dif}, $mon_{\_dif}$, $M_{cpltness}$) for determining the factors of concern for satellite stereo matching results. The methodological workflow is shown in ***Figure 3***, which consists of a series of geometric processing for generating needed statistics for analysis. For each stereo pair, we compute their $A_{unfilled}$ and A_{filled}, with which the other meta-parameters are used to correlate using a robust linear regression method. Geometric processing methods highlighted in gray in ***Figure 3*** will be briefly introduced in the following subsections (section 3.1 through 3.4).

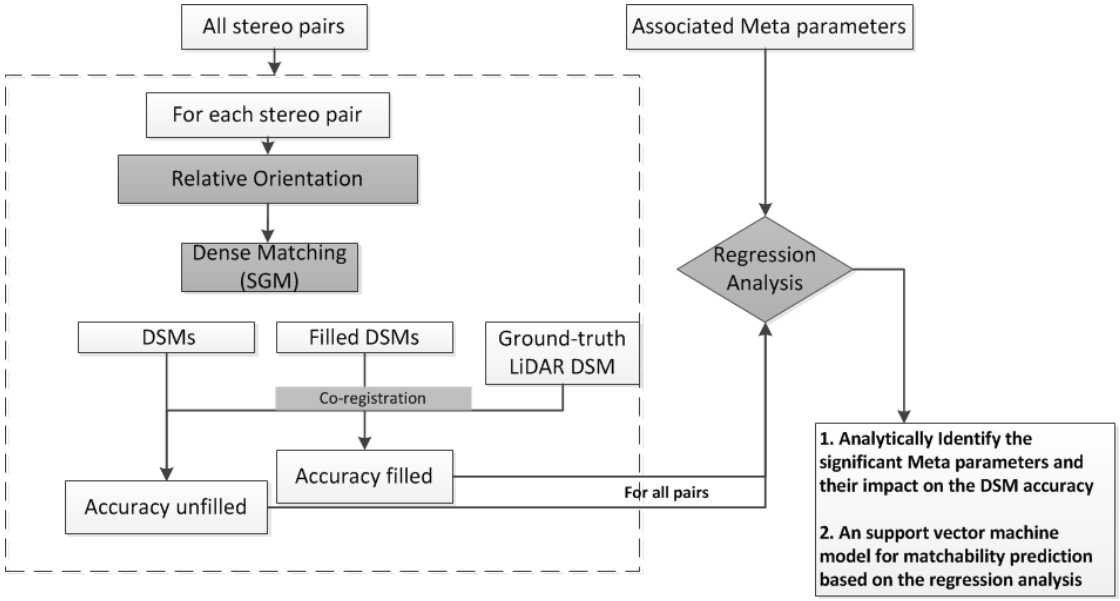


Figure 3. The methodological workflow of this study. Processing units highlighted in gray are introduced in subsections 3.1-3.4.

3.1. Relative Orientation

Most spaceborne imaging sensors are geometrically modeled through rational polynomial functions (RPF) (Fraser and Hanley, 2003), where the object-to-image mapping is constructed through a series of third-order polynomials in a non-linear fashion :

$$l = \Omega_l(U,V,W,\boldsymbol{a_1},\boldsymbol{b_1}) = \frac{L_{num}(U,V,W,\boldsymbol{a_1})}{L_{den}(U,V,W,\boldsymbol{b_1})}, \quad s = \Omega_s(U,V,W,\boldsymbol{a_1},\boldsymbol{b_1}) = \frac{S_{num}(U,V,W,\boldsymbol{a_2})}{S_{den}(U,V,W,\boldsymbol{b_2})} \tag{1}$$

where l and s are pixel locations representing **l**ine and **s**ample (y and x); U,V,W are the normalized object-space coordinates. For numerical purposes, these are normally linearly scaled to small values loosely between [-1, 1]. $\boldsymbol{a_i}$, $\boldsymbol{b_i}$, $i = 1,2$ are rational polynomial coefficients (RPC) for the four polynomial functions $L_{num}(\cdot)$,$L_{den}(\cdot)$, $S_{num}(\cdot)$ and $S_{den}(\cdot)$, computed through physical linear-array models. With each of $\boldsymbol{a_i}$, $\boldsymbol{b_i}$ a vector of 20 presenting 20 polynomial coefficients, it totals 80 parameters up to a scale

difference (for computing both l and s), thus 78. The relative orientation is carried out by a 0th order bias correction:

$$l + \Delta_l = \Omega_l(U, V, W, \boldsymbol{a_1}, \boldsymbol{b_1}),\ s + \Delta_s = \Omega_s(U, V, W, \boldsymbol{a_1}, \boldsymbol{b_1}) \quad (2)$$

Obviously, a single linear bias (Δ_l, Δ_s) might not be able to accommodate non-linear distortions in the raw satellite imagery (level-0 data) due to the inconsistent platform speed and rotations. Moreover, the follow-up epipolar rectification requires images to be at the same GSD to facilitate per-pixel dense matching. Therefore, prior to bias correction, a level-1 or level-2 geometric correction is needed that project the raw satellite images of a stereo to a common height plane to eliminate the non-systematic distortions between the scanning lines and at the same time resample the pixels to the same GSD. This involves recalculating the RPC parameters in the projected image:

$$[\widehat{\boldsymbol{a_1}}, \widehat{\boldsymbol{a_2}}, \widehat{\boldsymbol{b_1}}, \widehat{\boldsymbol{b_2}}] = \mathrm{argmin}_{\boldsymbol{a_1},\boldsymbol{a_2},\boldsymbol{b_1},\boldsymbol{b_2}} \sum_i |\Omega_l(U_i, V_i, W_i, \boldsymbol{a_1}, \boldsymbol{b_1}) - l_i|^2 + |\Omega_s(U_i, V_i, W_i, \boldsymbol{a_2}, \boldsymbol{b_2}) - s_i|^2 \quad (3)$$

where $\{U_i, V_i, W_i, l_i, s_i\}_i$ refer a set of generated points from the original RPC model through re-projecting virtual control points to the common height plane Π_h. These virtual control points are generated by assuming known regular gridded points in the space, and their image-space points are computed by firstly projecting them to the original image, and then project them back to Π_h, where the pixel locations of the virtual control points can be computed based on the GSD (ground sampling distance) of the projected image. Normally it is not necessary to project both images onto the same height plane for correction, while in our experiment this is for eliminating the scale differences for epipolar image rectifications.

Equation (3) for computing the new RPC parameters is a non-constrained minimization problem and can be solved iteratively using a classic Gauss-newton method (Boyd and Vandenberghe, 2004), with the initial values normally being the original RPC parameters $\boldsymbol{a_i}, \boldsymbol{b_i}, i = 1,2$. The refined RPC parameters $\widehat{\boldsymbol{a_1}}, \widehat{\boldsymbol{a_2}}, \widehat{\boldsymbol{b_1}}, \widehat{\boldsymbol{b_2}}$ and the level-1 image are used for the bias correction as in Equation (2).

The biases in Equation (2) to be computed theoretically require only one pair of corresponding points as the minimal solver. The goal is to adjust epipolar line of the point $\boldsymbol{p_1}$ in one view to cross the corresponding point $\boldsymbol{p_2}$ in the other view. It is well known that in linear array cameras this theoretically non-straight line (quasi-epipolar line) can be approximated as a straight (Morgan et al., 2004), especially for geometrically corrected product (level-1/2) of typically sized regions (a few hundreds of square kilometers). Thus the biases can be calculated as the vector compensating the misalignment between the epipolar line and corresponding point (Kuschk et al., 2014; Wang et al., 2011): for a pair of corresponding points $\boldsymbol{p_1}$, $\boldsymbol{p_2}$ respectively on image I_1 and I_2, with ℓ_{12} being the $\boldsymbol{p_1}$'s corresponding epipolar line, the biases can be simply computed as the vector from $\boldsymbol{p_2}$ to ℓ_{12}.

Given that the minimal solver requires only one pair of corresponding points, the blunders yielded by typical point feature matches (e.g. SIFT/SURF) (Bay et al., 2006; Lowe, 2004) can be effectively eliminated through a RANSAC (Random Sampling Consensus) estimator (Fischler and Bolles, 1981). This requires only a few iterations (5-10 iterations) even for a blunder rate of 50 percent (which in practice is much lower). Once the biases are computed, the epipolar rectification can be corrected through a simple rigid 2D image transformation to align paralleled epipolar lines, for more details the readers may refer to (Kuschk et al., 2014).

3.2. Core Matching Algorithm – Hierarchical Semi-Global Matching (SGM)

We use the Semi-global matching (SGM) algorithm as the core dense matching method for our evaluation, where we use the Census matching cost, as reported to be one of the most radiometrically robust metrics as was evaluated by Hirschmüller et al. (Hirschmüller, 2008). Although SGM is no longer the top algorithm in the lead-board of many benchmark tests (Scharstein and Szeliski, 2014), for the

moment it is still the top choice in the industry and many academic research studies, due to its well-leveraged memory demand, computational efficiency and implementation efforts. Very recent deep learning based works (Park and Yoon, 2018; Seki and Pollefeys, 2017; Zbontar and LeCun, 2016) are still based on the baseline SGM methods. Through our experiments, we attempted Convolutional Neural networks (CNN) (Zbontar and LeCun, 2016) based method, which in general yielded slightly better results while returning similar conclusions. Thus in this paper, our observations and findings are based on the SGM method with the use of traditional Census cost.

The implementation of the SGM is realized in a hierarchical manner in the author's prior software package RPC stereo processor (RSP) (Qin, 2016b), which is capable of efficiently processing large-format images with different types of image bit-depth. We used Census as the cost metric in the SGM optimization using a 7×9 window, and in our experiments we take the 16-bit panchromatic satellite images for matching. The implementation being hierarchical first builds an image pyramid, and then takes the computed disparity from coarser level as an initial for the finer level, where the disparity range in the finer level is dynamically defined based on the disparity from the coarser level **(**Rothermel et al., 2012**)**. This simple strategy can dramatically reduce memory demand and computation time, and the number of pyramids is dynamically determined based on the computer memory and the pre-defined disparity range, which is tunable in the RSP hyper-parameter set; we use [-1000,1000] which is more than enough in our test. For more details about SGM and our particular implementations, the readers may refer to (Hirschmüller, 2005; Qin, 2014; Qin, 2016a).

3.3. DSM co-registration

Since we only perform a relative orientation for each of the pairs, there exists misalignment of the resulting DSMs and the LiDAR DSM. Thus before evaluating the accuracies of the DSMs, co-registration between the generated DSMs and the LiDAR DSM is needed. It is noted that this is a desired process over a bundle adjustment (BA) for all the images using LiDAR DSM as where the control points are selected, as our goal is to evaluate stereo pairs, thus 1) relative orientation ensures the orientation of one pair is not affected by inaccurate feature measurements of the other pairs; 2) independent co-registration of DSMs from each pair with the LiDAR DSM ensures a fair comparison that only evaluates the fit of a surface by compensating systematics errors.

Given that in the top-view dataset from satellite images, the rotation differences in the co-registration are normally ignorable (Waser et al., 2008), and also co-registering these many DSMs requires consideration of computational simplicity, our co-registration process is rather intuitive: an adaptive registration process is performed considering only translation differences. Since the computed DSMs in our evaluation dataset is accurate in the range of 0 – 15 meters in all direction (given the absolution positional errors of Worldview-3 dataset), the adaptive registration method searches the translation through a dichotomy division. To ensure the blunders not affecting the registration, we discard points that are larger than 6 meters (empirical value through the standard deviation of the DSM) in the registration processes.

3.4. Robust linear-regression method

The approximately 1,180 DSM products from the combinatorial stereo pairs yield data points in terms of $A_{unfilled}$ and A_{filled} corresponding to different meta-parameters (***Table 1***)**.** A few of the stereo pairs failed to generate DSMs or only generating very sparse and unreliable point clouds for reasons including 1) too large radiometric differences and 2) too large/too small of intersection angles. We also noticed the systematic errors of some of the resulting DSMs exceeded the 15-meter threshold, resulting in very large $A_{unfilled}$ and A_{filled} values. We tend to keep our processing procedure standard and regarded these data points as outliers given that we have sufficient data points for analysis. Therefore considering

these outliners, we use a robust linear regression method that reweights iteratively the observations based on their residuals to for solving the regression problem (Huber, 2011; Street et al., 1988), where the weights are computed through a kernel mapping:

$$\boldsymbol{w} = T[|\boldsymbol{b}| < \mathbf{1}] \cdot (1 - \boldsymbol{b}^2)^2, \quad \boldsymbol{b} = \boldsymbol{r}/(t \cdot s \cdot \sqrt{(\mathbf{1} - \boldsymbol{h})}) \tag{4}$$

where bold letters and numbers refer to vectors, and all the operations on the vectors in Equation (4) are element-wise. $\boldsymbol{b}$ is a normalized residual vector, and the normalization term $(t \cdot s \cdot \sqrt{(\mathbf{1} - \boldsymbol{h})}$ is element-wise normalization value per iteration that considers a user-defined tuning constant t, an estimate of standard derivation s through the means of maximal mean derivation, and a per point uncertainty estimates in the last iteration through the diagonal elements of the covariance matrix (leverage score $\boldsymbol{h}$) . We took all the values from a standard MATLAB implementation, and for details the readers may refer to (Mathworks, 2018). The weight $\boldsymbol{w}$ tends to eliminate data points with normalized residual $\boldsymbol{b}$ larger than 1, and gives higher weight for those with smaller residuals. Results of our analytical results are shown in Section 4.

4. Analytical Results

For the three test regions in ***Figure 2***, we compute the correlation between the meta-parameters summarized in ***Table 1*** and the DSM accuracy. There are in total eight meta-parameters for the assessment, and in total forty-eight regression results with respect to the accuracy of the filled and unfilled DSM in the three test regions are performed. It is expected that among these eight parameters, some of them may be inter-correlated and the relatively significant parameters should be identified for analysis, for example, based on Table 1, the Sun angle difference θ_{s_dif} computed as a summed squared root of $\theta_{s\_a\_dif}$ and $\theta_{s\_e\_dif}$, and is naturally correlated with them.

4.1. Significant meta-parameters

The most straightforward way to identify the correlated parameters is to analyze the covariance matrix of the meta-parameters and the accuracy of the DSM. However, this does not offer enough cues to their roles in contributing to the potential achievable accuracy, for example, their combinatorial contributions. Hence, we use a typical statistical measure – coefficient of determination (R^2) in the regression analysis to understand the most significant parameters: we enumerate all possible combinations of the parameter set of these eight meta-parameters and measure their $\boldsymbol{R^2}$ values of linear regressions. $\boldsymbol{R^2}$ in a regression analysis generally refers to the degree of fit of the model to the data points, and is defined as follows:

$$\begin{cases} R^2 = 1 - \frac{R_{res}}{R_{tot}} \\ R_{res} = \sum_i (y_i - f_i)^2 \\ R_{tot} = \sum_i (y_i - \bar{y})^2 \end{cases} \tag{5}$$

where R_{res} refers to the residuals of fit and R_{tot} refers to the variance of the data; y_i and $\bar{y}$ refer to the observations and their mean, and f_i refer to the predicted values of the regression model. In general, the $\boldsymbol{R^2}$ values indicate the amount of data variance accounted (or explained) by the variables (i.e. meta-parameters in our context) through the regression model, meaning that, the larger $\boldsymbol{R^2}$ value, the better accuracy prediction the variables can be used for, and in our contexts variables refer to the meta-parameters and the predictions refer to the DSM accuracy. In Table 2 we show the $\boldsymbol{R^2}$ values of all possible combinations of the parameters with respect to the accuracy of the filled DSM. The rationale for evaluating the filled DSM is because it is the final geometric product used in practice. To save space, we list parameter sets with the top-3 $\boldsymbol{R^2}$ values under each scenario (# parameters used for regression).

Table 2. $\boldsymbol{R^2}$ values of regression analysis using different meta-parameter set (top-3 for each scenario)

#	Region 1		Region 2		Region 3	
	Parameters used	$\boldsymbol{R^2}$	Parameters used	$\boldsymbol{R^2}$	Parameters used	$\boldsymbol{R^2}$

1	$[\theta_{s\_dif}]$	0.301013	$[\theta_{s\_dif}]$	0.252847	$[\theta_{s\_dif}]$	0.351888
	$[\theta_{s\_e\_dif}]$	0.298236	$[\theta_{s\_e\_dif}]$	0.247662	$[\theta_{max\_n}]$	0.221411
	$[\theta_{s\_a\_dif}]$	0.244504	$[\theta_{s\_a\_dif}]$	0.201118	$[\theta_{s\_e\_dif}]$	-0.089758
2	$[\theta_{in}, M_{cpltness}]$	0.417887	$[\theta_{in}, M_{cpltness}]$	0.425199	$[\theta_{s\_e\_dif}, mon_{\_dif}]$	0.443441
	$[\theta_{in}, \theta_{s\_dif}]$	0.335602	$[\theta_{max\_n}, \theta_{s\_dif}]$	0.263291	$[\theta_{s\_e\_dif}, \theta_{s\_dif}]$	0.439759
	$[\theta_{in}, \theta_{s\_e\_dif}]$	0.333289	$[\theta_{s\_dif}, mon_{\_dif}]$	0.262187	$[\theta_{s\_dif}, mon_{\_dif}]$	0.426932
3	$[\theta_{in}, \theta_{s\_a\_dif}, M_{cpltness}]$	0.422444	$[\theta_{in}, \theta_{s\_a\_dif}, M_{cpltness}]$	0.42683	$[\theta_{in}, \theta_{max\_n}, M_{cpltness}]$	0.449761
	$[\theta_{in}, \theta_{max\_n}, M_{cpltness}]$	0.421934	$[\theta_{in}, \theta_{s\_e\_min}, M_{cpltness}]$	0.426705	$[\theta_{s\_e\_min}, \theta_{s\_e\_dif}, \theta_{s\_a\_dif}]$	0.448370
	$[\theta_{in}, mon_{\_dif}, M_{cpltness}]$	0.421489	$[\theta_{in}, mon_{\_dif}, M_{cpltness}]$	0.423945	$[\theta_{s\_e\_min}, \theta_{s\_dif}, mon_{\_dif}]$	0.447775
4	$[\theta_{in}, \theta_{max\_n}, mon_{\_dif}, M_{cpltness}]$	0.425502	$[\theta_{in}, \theta_{s\_e\_min}, \theta_{s\_a\_dif}, M_{cpltness}]$	0.427741	$[\theta_{in}, \theta_{s\_e\_dif}, \theta_{s\_a\_dif}, M_{cpltness}]$	0.451880
	$[\theta_{in}, \theta_{s\_a\_dif}, mon_{\_dif}, M_{cpltness}]$	0.424179	$[\theta_{in}, \theta_{s\_a\_dif}, mon_{\_dif}, M_{cpltness}]$	0.424071	$[\theta_{in}, \theta_{s\_e\_min}, \theta_{s\_dif}, M_{cpltness}]$	0.451880
	$[\theta_{in}, \theta_{max\_n}, \theta_{s\_a\_dif}, M_{cpltness}]$	0.423552	$[\theta_{in}, \theta_{max\_n}, \theta_{s\_e\_min}, M_{cpltness}]$	0.423225	$[\theta_{in}, \theta_{max\_n}, mon_{\_dif}, M_{cpltness}]$	0.451880
5	$[\theta_{in}, \theta_{max\_n}, \theta_{s\_a\_dif}, mon_{\_dif}, M_{cpltness}]$	0.425739	$[\theta_{in}, \theta_{max\_n}, \theta_{s\_e\_min}, \theta_{s\_a\_dif}, M_{cpltness}]$	0.423120	$[\theta_{in}, \theta_{max\_n}, \theta_{s\_e\_dif}, mon_{\_dif}, M_{cpltness}]$	0.457555
	$[\theta_{in}, \theta_{max\_n}, \theta_{s\_dif}, mon_{\_dif}, M_{cpltness}]$	0.423297	$[\theta_{in}, \theta_{s\_e\_min}, \theta_{s\_a\_dif}, mon_{\_dif}, M_{cpltness}]$	0.421106	$[\theta_{in}, \theta_{max\_n}, \theta_{s\_e\_dif}, \theta_{s\_a\_dif}, M_{cpltness}]$	0.453843
	$[\theta_{in}, \theta_{max\_n}, \theta_{s\_e\_min}, mon_{\_dif}, M_{cpltness}]$	0.422395	$[\theta_{in}, \theta_{max\_n}, \theta_{s\_a\_dif}, mon_{\_dif}, M_{cpltness}]$	0.420123	$[\theta_{in}, \theta_{max\_n}, \theta_{s\_a\_min}, \theta_{s\_e\_dif}, M_{cpltness}]$	0.453843
6	$[\theta_{in}, \theta_{max\_n}, \theta_{s\_e\_min}, \theta_{s\_a\_dif}, mon_{\_dif}, M_{cpltness}]$	0.421619	$[\theta_{in}, \theta_{max\_n}, \theta_{s\_e\_min}, \theta_{s\_a\_dif}, mon_{\_dif}, M_{cpltness}]$	0.417374	$[\theta_{in}, \theta_{max\_n}, \theta_{s\_e\_min}, \theta_{s\_e\_dif}, mon_{\_dif}, M_{cpltness}]$	0.456597
	$[\theta_{in}, \theta_{max\_n}, \theta_{s\_e\_min}, \theta_{s\_dif}, mon_{\_dif}, M_{cpltness}]$	0.419715	$[\theta_{in}, \theta_{max\_n}, \theta_{s\_e\_min}, \theta_{s\_e\_dif}, \theta_{s\_dif}, M_{cpltness}]$	0.410811	$[\theta_{in}, \theta_{max\_n}, \theta_{s\_e\_dif}, \theta_{s\_dif}, mon_{\_dif}, M_{cpltness}]$	0.456597
	$[\theta_{in}, \theta_{max\_n}, \theta_{s\_a\_dif}, \theta_{s\_dif}, mon_{\_dif}, M_{cpltness}]$	0.414960	$[\theta_{in}, \theta_{max\_n}, \theta_{s\_e\_min}, \theta_{s\_e\_dif}, \theta_{s\_a\_dif}, M_{cpltness}]$	0.410811	$[\theta_{in}, \theta_{max\_n}, \theta_{s\_e\_dif}, \theta_{s\_a\_dif}, mon_{\_dif}, M_{cpltness}]$	0.456597
7	$[\theta_{in}, \theta_{max\_n}, \theta_{s\_e\_dif}, \theta_{s\_a\_dif}, \theta_{s\_dif}, mon_{\_dif}, M_{cpltness}]$	0.414960	$[\theta_{in}, \theta_{max\_n}, \theta_{s\_e\_min}, \theta_{s\_e\_dif}, \theta_{s\_a\_dif}, \theta_{s\_dif}, M_{cpltness}]$	0.410811	$[\theta_{in}, \theta_{max\_n}, \theta_{s\_e\_dif}, \theta_{s\_a\_dif}, \theta_{s\_dif}, mon_{\_dif}, M_{cpltness}]$	0.456597
	$[\theta_{in}, \theta_{max\_n}, \theta_{s\_e\_min}, \theta_{s\_e\_dif}, \theta_{s\_a\_dif}, mon_{\_dif}, M_{cpltness}]$	0.414003	$[\theta_{in}, \theta_{max\_n}, \theta_{s\_e\_min}, \theta_{s\_e\_dif}, \theta_{s\_dif}, mon_{\_dif}, M_{cpltness}]$	0.403003	$[\theta_{in}, \theta_{max\_n}, \theta_{s\_e\_min}, \theta_{s\_a\_dif}, \theta_{s\_dif}, mon_{\_dif}, M_{cpltness}]$	0.454271
	$[\theta_{in}, \theta_{max\_n}, \theta_{s\_e\_min}, \theta_{s\_a\_dif}, \theta_{s\_dif}, mon_{\_dif}, M_{cpltness}]$	0.414003	$[\theta_{in}, \theta_{max\_n}, \theta_{s\_e\_min}, \theta_{s\_a\_dif}, \theta_{s\_dif}, mon_{\_dif}, M_{cpltness}]$	0.403003	$[\theta_{in}, \theta_{s\_e\_min}, \theta_{s\_e\_dif}, \theta_{s\_a\_dif}, \theta_{s\_dif}, mon_{\_dif}, M_{cpltness}]$	0.454271
8	$[\theta_{in}, \theta_{max\_n}, \theta_{s\_e\_min}, \theta_{s\_e\_dif}, \theta_{s\_a\_dif}, \theta_{s\_dif}, mon_{\_dif}, M_{cpltness}]$	0.414003	$[\theta_{in}, \theta_{max\_n}, \theta_{s\_e\_min}, \theta_{s\_e\_dif}, \theta_{s\_a\_dif}, \theta_{s\_dif}, mon_{\_dif}, M_{cpltness}]$	0.403003	$[\theta_{in}, \theta_{max\_n}, \theta_{s\_e\_min}, \theta_{s\_e\_dif}, \theta_{s\_a\_dif}, \theta_{s\_dif}, mon_{\_dif}, M_{cpltness}]$	0.454271

Prior to the R^2 computation, we pre-filter the data points using a 3-sigma rule to eliminate the blunders due to DSM mis-registration. Table 2 shows the meta-parameters that achieves the leading R^2. It can be seen that when all the eight parameters were used for regression, it achieves R^2 of ca. 0.4-0.45 through all the three regions, meaning that approximately 40% - 45% of the uncertainty of A_{filled} can be accounted/predicted using the meta-parameters. It should be noted that such values can be low in a typical statistical analysis; while given the complexity in terms of imaging conditions, image matching algorithms and scene contents, predicting accurately the resulting DSM accuracy merely using a few meta-parameters are technically not achievable, while with these uncertainties in a simple linear regression accounted solely by meta-parameters can be already very valuable. A few observations are: first of all, although the three regions refer to respectively different scene contexts, the values 0.4-0.45 of R^2 reflect that the statistics are fairly consistent and not much dependent on the regions; secondly, this amount of uncertainty could be readily accounted using fewer parameters instead of all. For example, in row #3, region 1, simply using " $[\theta_{in}, \theta_{s\_a\_dif}, M_{cpltness}]$" achieves similar and sometimes even higher R^2 values than using all the meta-parameters, and this is consistently observed in other regions as well. This indicates that there exist inter-correlations among these meta-parameters and three of them are sufficient. Meanwhile, we also observe that in table 2 (Region 1 and 2) some of the R^2 values using fewer parameters are even slightly higher than those using more. We understand in this case, more parameters, especially those not linearly correlated with the resulting accuracy, might introduce errors and randomness leading to lower R^2 values. Among these top-3 parameters, the most frequently observed parameters for scenarios with three parameters or less (row # 1-3) are respectively the intersection angle θ_{in}, sun angle differences θ_{s_dif} and the matching completeness $M_{cpltness}$. These two angles are all pre-matching meta-parameters and can be computed directly from the satellite geometric parameters. Note that the θ_{s_dif} is the summed squared roots of the sun elevation $\theta_{s\_e\_dif}$ and

azimuth $\theta_{s\_a\_dif}$ angle, and based on their appearing frequency on the table (four for $\theta_{s\_e\_dif}$ and five for$\theta_{s_a_dif}$ in row # 1-3), there is no clear clue that one is more significant over the other. $M_{cpltness}$ is a post-matching parameter and can be dependent on different matching algorithms. Being said, to have a quick and qualitative assessment on the quality of a stereo pair for DIM, $\theta_{in}$and θ_{s_dif} can be effective.

4.2. Understanding the associations between the meta-parameters and the DSM accuracy

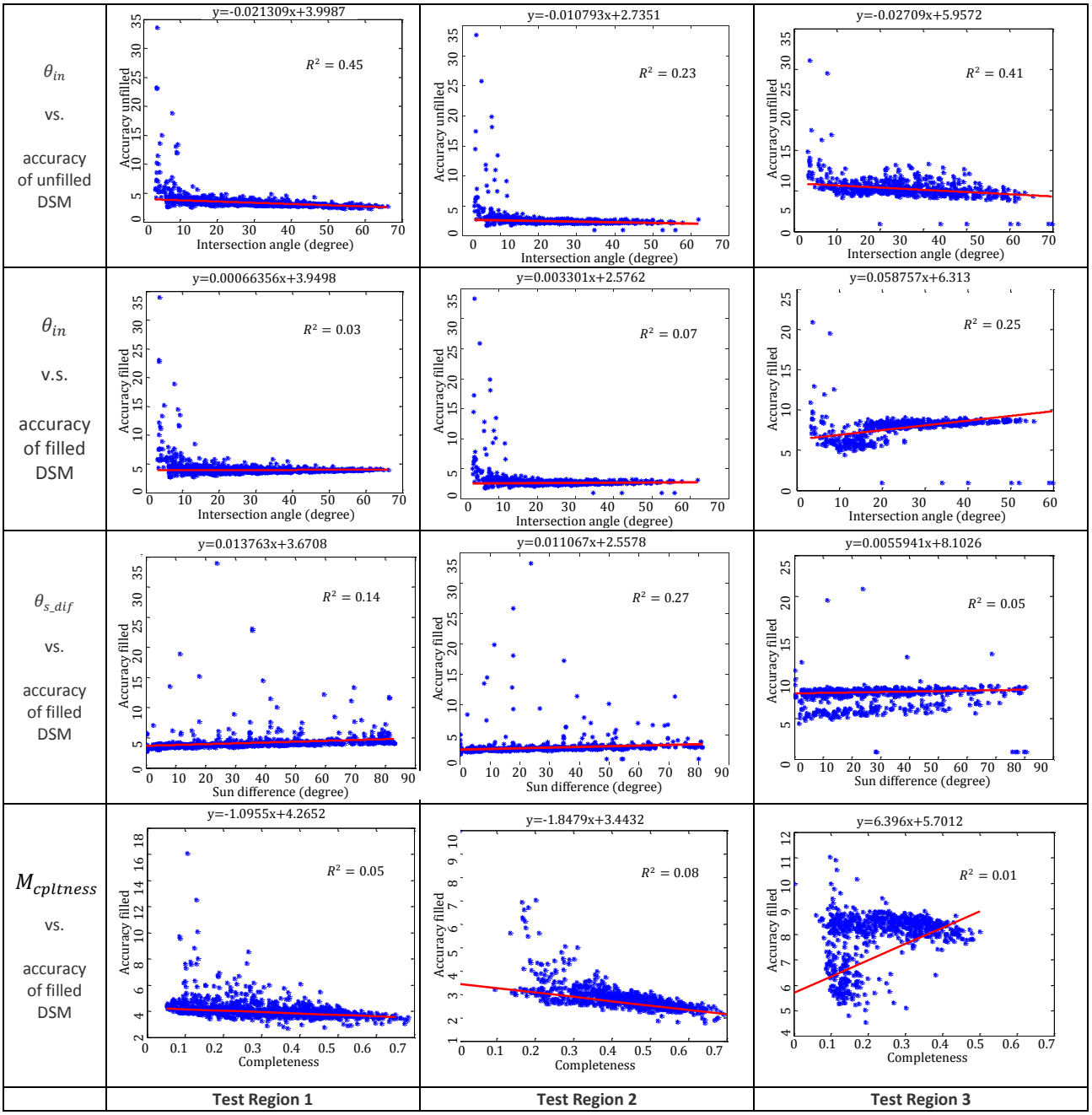


Figure 4. The regression analysis for intersection angle θ_{in}, maximal off-nadir angle θ_{max_n} and sun angle difference θ_{s_dif}. These very small R^2 (<0.1) values show that the variation of the data is not well accounted by a single meta-parameter.

Figure 4 plots the single-variable linear regression of these three parameters (θ_{in}, θ_{s_dif} and $M_{cpltness}$) with respect to the accuracy of the DSM (filled or unfilled, selectively) of three test regions. Although a

single parameter might not well account the uncertainties of the data (very low R^2 values), it provides cues to understand how each of these three parameters is correlated (positive/negative). In particularly we plot the regression line of θ_{in} with respect to both the unfilled and filled DSM accuracy due to observed differences, while for the other two parameters (i.e., θ_{s_dif} and $M_{cpltness}$), we only plot the regression line in ***Figure 4*** with respect to the filled DSM accuracy as those with respect to unfilled DSM accuracy are similar. Each data point in a subplot refers to the accuracy (RMSE) of a DSM generated from a stereo pair. It can be seen that there exist a few data points with very large error values (e.g. larger than 10 meters), and most of these are blunders due to misregistration error (systematic errors in the Z-direction exceeds the pre-defined 15-meter search range). Through ***Figure 4*** it can be seen that with robust linear-regression these blunders can be well accounted.

A. Intersection angle

Among these three test regions, the correlation between the intersection angle and the unfilled DSM clearly indicates that the larger intersection angle, the higher measurement error. This is in line with our understanding on the association of intersection angles with per-point accuracy: a larger intersection angle indicates a larger base-high ratio resulting in better vertical accuracy given the same amount of horizontal uncertainties (also evidenced by the large R^2 value). However, a stereo pair with larger intersection angles may introduce the more occlusions, thus yielding high accuracy but incomplete DSM. This is shown in the second row of ***Figure 4***: the correlation between the intersection angle and the filled DSM demonstrates a different trend, being that a smaller intersection angle in general yields better results for the filled DSM (lower RMSE). This at a first glance, seems to be a contradictory conclusion to the case when using the unfilled DSM for evaluation, while in this case the filled DSM indeed taking the completeness of the DSMs into consideration: although a larger intersection angle theoretically offers higher accuracy in terms of per-point measurement, the occluded regions being interpolated in overall brings negative impacts as the intersection angle becomes larger. This becomes particularly critical for stereo matching on scenes with high-rise buildings, evidenced by test region 3 (see ***Figure 4***, first row, fourth column): in its scatter plot, there is an obvious discontinuity at the point of 18 degrees, from where onwards the error becomes particularly high. This is because that the high-rise buildings create large parallax leading to large occlusions and non-matched areas, which brings down the accuracy significantly, thus in such a case smaller intersection angles is beneficial. As a result, in general, we observed that a small intersection angle (can be as small as around 7°) may generate visually and statistically reasonable results for the filled DSMs.

B. Sun angles

As shown in the third row of ***Figure 4***, we also observed that the difference between the Sun angles θ_{s_dif} for two images of a stereo pair plays an important role that impacts on the accuracy of the generated DSM (an example shown in ***Figure 1***). From the regression line in ***Figure 4*** (row #3), it clearly indicates that the resulting filled DSM accuracy is highly correlated with the Sun angle difference θ_{s_dif}, and the larger θ_{s_dif} is, the larger the RMSE of the associated DSMs. The rationale behind this can be interpreted: since the ground surface are much more complicated than a direction-free Lambertian reflection, and can be a mixture of complex specular and anisotropic surfaces (Szeliski, 2010), a slight lighting condition difference (i.e. different Sun angles) might result in completely different textures in certain regions, thus leading to matching failures. Its impact, if comparing to the intersection angles (row #2 of ***Figure 4***) is even more significant (the fitted line has a steeper slope as well as higher R^2) on the filled DSM accuracies. This necessarily indicates that this parameter should be critically considered when forming incidental (or cross-track) stereo pairs.

C. Matching completeness

The completeness is a post-matching parameter that evaluates the percentage of successfully matched pixels in the DSM (not necessarily correct). The unmatched pixels are normally eliminated through a widely used (and standard) process called "left-right consistency check", and this process is also effectively used for occlusion detections (Kolmogorov and Zabih, 2001). As a result, the unmatched regions can be interpolated using standard methods (e.g. Inversed distance weight (Lu and Wong, 2008), triangulation-based (Jordan, 2007), or membrane fitting methods (Pérez et al., 2003)). As of our data, we used Delaunay triangulations for interpolation (Qin, 2016a). It is expected the interpolated areas may often return large errors than the directly matched pixels, therefore the rationale for $M_{cpltness}$ is that, if the percentage of the matched pixels are higher (larger $M_{cpltness}$), only a small percentage of pixels needs to be interpolated thus the filled DSM should return higher accuracy. The presented regression line in ***Figure 4*** (row #4) shows such a correlation in the first two regions, while in regions 3, such correlation is weak; this may be due to other factors affecting on the accuracy of the matched pixels, for example, the due to high-rise buildings in the scene introduce more occlusions and terrain distortions (in the epipolar space) affecting the matching, etc. Therefore, such correlation can be positive while as shown across three regions, can be scene dependent. In addition, as a post-matching parameter, its uses for predicting potential generated DSM quality can be dependent on algorithms as well, and of course, it is infeasible to compute the association of such parameter for every single algorithm, whereas it still can be indicative and valuable for typical/standard algorithms (as in our case we used SGM).

5. A Support Vector Machine Model for Stereo Pair Quality Prediction with application to Multi-view Stereo 3D Reconstruction

Being able to predict the quality of a stereo pair can benefit domain experts for selecting good pairs for relevant applications, and automate the processes for multi-view stereo (MVS) 3D reconstruction. An often used strategy for reconstructing DSM from MVS images is to fuse individual DSMs generated from stereo pairs (d'Angelo and Kuschk, 2012; Qin, 2017). Theoretically, the fused DSM quality will improve as the number of individual DSMs increases assuming the expectation of DSMs being unbiased, while this can be impractical both in terms of image availability and computational load. Here based on the data used in our previous analysis, we train a support vector machine model that predicts the level of quality for stereo pairs, and take this model as a pair selector to achieve fully automated MVS 3D reconstruction.

5.1. A trained support vector machine model for stereo pair quality prediction

An intuitive idea is to take a linear regression curve used in ***Section 4*** to predict the DSM accuracy of the stereo pair. This is however problematic given the low $\boldsymbol{R^2}$ scores (***Table 2***). This is not well generalized as data points are all from worldview 2/3 data with a GSD of 0.3-0.5 meters. Essentially, a generalized quality indicator should be unit-less and thus can be applied to images from different sensors and resolutions. Hence we consider categorizing our data points (metadata of stereo pairs in and the resulting DSM accuracy (filled) as described ***Table 1***) based on their DSM accuracies. For all data points in test Region 1, DSM accuracy ranges from 2.7 to 16 meters. This apparently contains blunders possibly due to registration errors. Normally as the accuracy is as low as 4 meters, the resulting DSMs are visually unacceptable (examples can be found in ***Figure 1.***). We hence categorize all data points in test region 1 with four discrete quality scores QS as follows:

$$QS = \begin{cases} 1 & A_{filled} \in [0,3)\ m \\ 2 & A_{filled} \in [3,3.5)\ m \\ 3 & A_{filled} \in [3.5,4)\ m \\ 4 & A_{filled} \in [4,+\infty)\ m \end{cases} \quad (6)$$

These discrete QS can be taken as training labels for the meta-parameters. The [meta-parameters, QS] pairs can be fed into a statistical classifier (e.g. Support Vector Machine (SVM) (Wang, 2005)) for training and the trained model can then be used for predicting the quality of a stereo pair. In our experiment, we simply take the meta-parameters listed in ***Table 1*** as the feature vector for training. Note we in our experiment we did not incorporate the completeness value $M_{cpltness}$, since it is a post-matching parameter are very often unavailable prior to computation. In our experiment, we use the RBF (radial-basis function) as the kernel function for training. Since SVM is well established, the readers may refer details to relevant materials (Foody and Mathur, 2004; Pal and Mather, 2005; Wang, 2005). To generate continuous QS cores, we simply interpolate for each predicted label using its neighboring labels through the confidence score (between 0 and 1; in SVM, this is provided as the normalized distance to the support vectors (Jiang et al., 2008; Wang, 2005)), being:

$$QS = L_{pd-1} * C_{pd-1} + L_{pd} * C_{pd} + L_{pd-1} * C_{pd-1} + L_{pd+1} * C_{pd+1} \quad (7)$$

where "pd" refers to the predicted label and "$pd-1$", "$pd+1$" refer to its neighboring labels. L and C refer to the label value itself (here in our case are any of 1,2,3,4) and the confidence level. Note the three C are normalized to 1 to ensure the QS within the range. Through our experiments, we find our strategy of categorizing the A_{filled} to discrete quality levels appears to be more robust than using regressions (e.g. support vector regression (Smola and Schölkopf, 2004)), as the model can be easily affected by data noises.

5.2. Experiment setup and results

We take the trained SVM model on test region 1 as described in 5.1., and use the quality level indicator to rank the ca. 1,180 stereo pairs with their predicted DSM accuracy (from high accuracy to low). The DSMs of the first five pairs (empirically determined based on an analysis in (Qin, 2017)) will be registered and processed through a median filter to produce the final DSM, and this process is shown in ***Figure 5***. The test will be performed on region 2 and 3 (***Figure 2***). It may be arguable that we test the model on regions covered by the same set of images. However, we observe that a DSM from a stereo pair that achieves the best accuracy in one region does not necessarily obtain the best accuracy on other regions, meaning that uncertainties can be spatially variant, thus it is essential to test if the model learned from one region can well-generalize the quality predictability over different regions. In addition, to test the transferability of the trained model, we have also performed experiments on a different MVS dataset (Jacksonville dataset) provided by the SpaceNet Challenge (SpaceNet, 2018).

As a comparison, we evaluated three other different ranking methods: 1) Example-based ranking that simply ranks the pairs based on the filled DSM accuracy of test region 1 and reuses the rankings to other regions; this method is obviously limited in scalability but can here serve as a good reference. 2) Intersection-angle based selection criterion (d'Angelo et al., 2014) that randomly selects stereo pairs whose intersection angles that are within 15-25°; 3) an intuitive ranking method based on (Facciolo et al., 2017): firstly pre-select pairs that meet certain criterion (i.e., intersection angle with 5-45°, and maximal off-nadir angle below 40°), and then rank the selected pairs based on their acquisition time interval (shorter the higher rank). DSMs from the top-five pairs out of each of these ranking will be fused through our workflow (***Figure 5***, highlighted red texts), and their accuracy will be evaluated against the ground-truth LiDAR data. The Median filter acts across the space the DSM stacking direction (concept shown in ***Figure 5***) and the outputting value for each pixel will be the median of the window cube: with a window size of 7×7 pixels the window cube has $7 \times 7 \times 5$ pixels, where "5" is the number of DSMs to be fused.

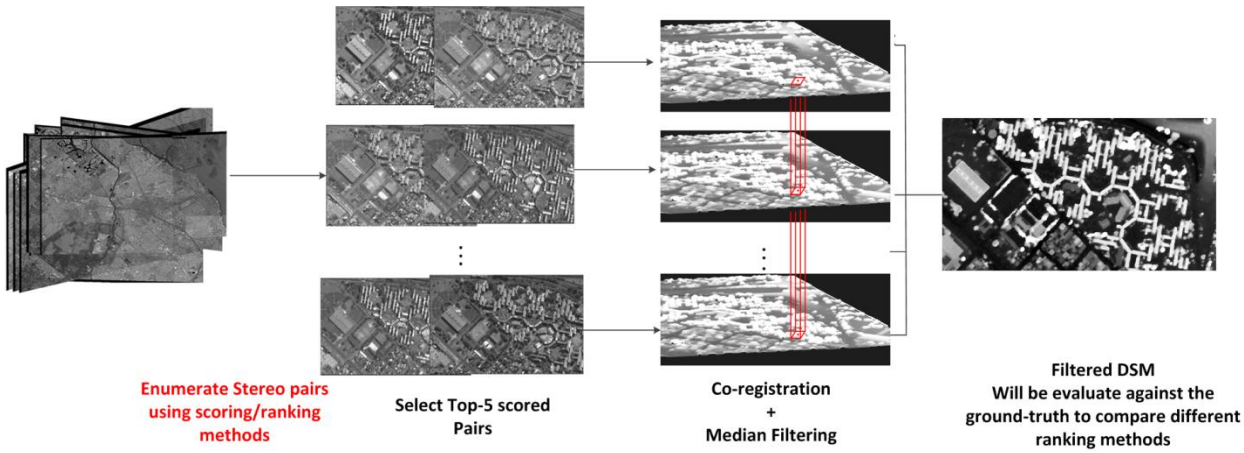


Figure 5. Workflow for DSMs fusion for Multi-stereo DSM generation. Different ranking methods including 1) Example-based ranking, 2) Intersection-angle criterion and our model 1 and model 2 regression formula (Equation (6) and (7)).

Table 3. Root-mean-squared error (RMSE) of MVS reconstruction using different pair selection methods (unit: meter)

Pair selection methods	***Test Region 2***	***Test Region 3***	***Jacksonville Region***
Example-based Pair Selection	1.17	1.86	N.A.
Intersection-angle based Selection (d'Angelo et al., 2014)	1.44	2.26	1.30
(Facciolo et al., 2017)	1.32	2.17	1.21
The proposed SVM-based Quality Indicator	**1.16**	**1.82**	**0.81**

Results of the MVS 3D reconstruction using different pair selection methods are shown in ***Table 3***. Different pair selection methods determine different top-five pairs used for 3D reconstruction using pipeline shown in ***Figure 5.*** The example-based pair selection method assumes at least a sub-region covered by the MVS dataset has ground-truth data, such that the quality of the pairs can be directly computed as the accuracy of that particular region. Given that in our experiment we use ground-truth data of test region 1 for example-based selection, the ranked pairs can only be reused in test region 2 and 3 as they share the same set of images. Intuitively, the example-based method should be the most competitive since the quality is directly evaluated based on the accuracy, while the SVM model surprisingly has a very similar performance as the example-based method. The intersection-angle based selection method performs worst among the four (ca. 0.3-0.5 meters higher in RMSE), and the method of (Facciolo et al., 2017) achieves better results but still significantly worse than Example-based method and our SVM-model based method. In the Jacksonville region, the SVM model pair selection method yields a much better result (0.49 meters smaller in RMSE) as compared to the intersection-based method and the method of (Facciolo et al., 2017). Visual results of the 3D reconstruction of the tested regions are shown in ***Figure 6:*** In each of the sub-figure, we have cropped an enlarged figure from an oblique view to show the details. For test region 2 and 3, the results from the example-based and SVM-based methods are highly similar. Indeed, we find that both methods selected the same top-five pairs for these two regions, with slight differences in their ranking sequence within the top five. Therefore, the subtle differences of the resulting DSM (max. 0.04m) are ignorable. However, the visual results of the intersection-angle based method and the method of (Facciolo et al., 2017) are notably worse in all three regions as compared to the other two methods. In the Jacksonville region, the SVM model based method show a notably better result both visually and quantitatively, even though the model is trained using information from a completely different dataset.

The results suggest two facts: 1) the proposed SVM-based matching quality indicator has the capacity to generalize the quality of stereo pairs and works equivalently well as example-based method. 2) The proposed quality indicator encapsulates different dimension of the stereo pair metadata and can serve as a general stereo-pair quality indicator and performs notably better than a simple intersection-angle based selection criterion and the method of (Facciolo et al., 2017)

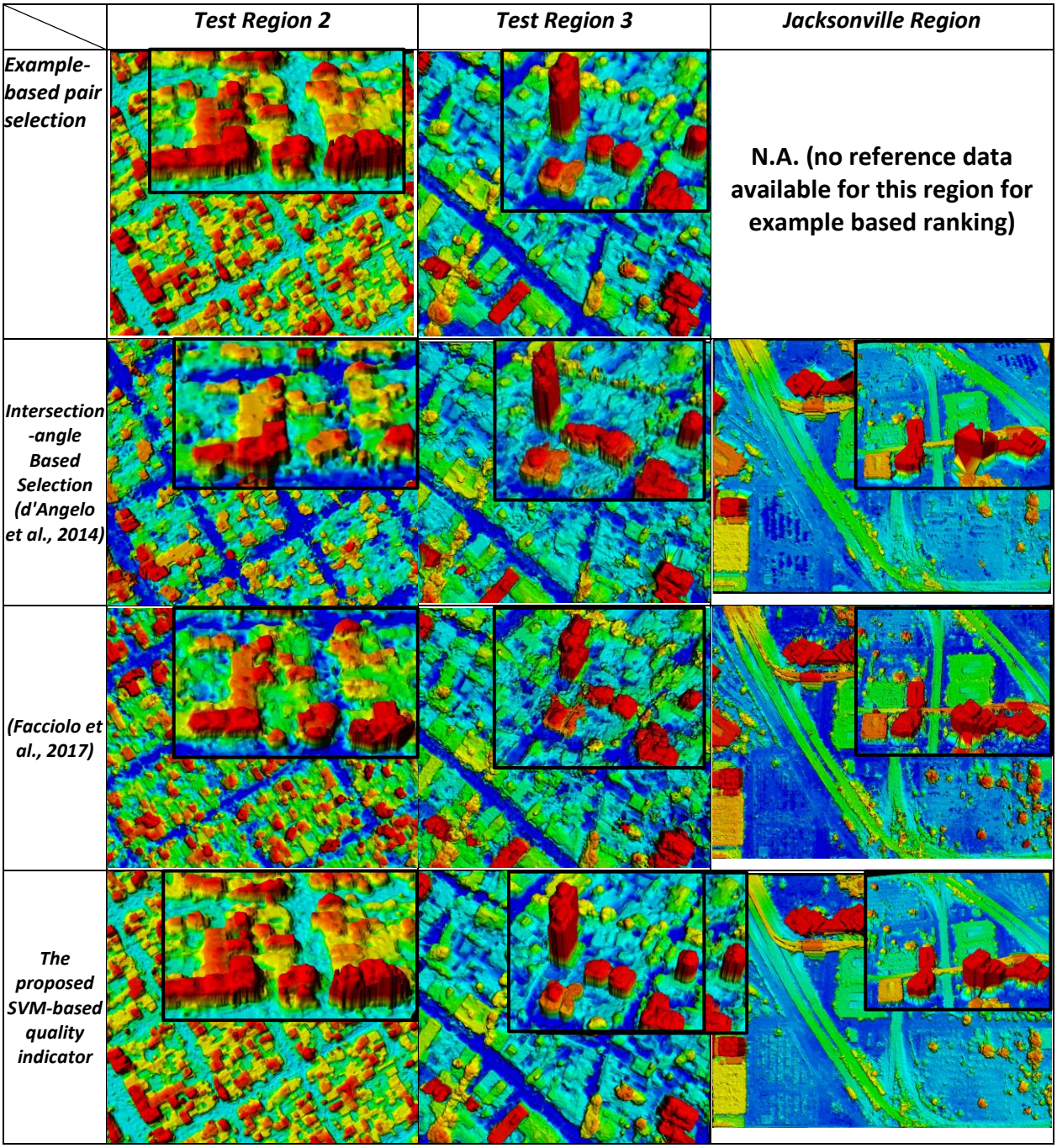


Figure 6. Visual results of the MVS reconstructed DSM from top-five pairs selected from different pair selection methods. The DSM is independently color-coded by height to optimize the visualization.

6. Discussions and Conclusions

This paper analyzes the relationship between critical meta-parameters (***Table. 1***) of satellite stereo pairs and their resulting accuracy, and has proposed a method that uses SVM model for training a general stereo pair quality indicator. DSMs generated from approximately 1,180 stereo pairs cross three typical urban regions have been evaluated against ground-truth, which serve as data points for analysis. An SVM-based quality indicator trained using data points from one of the regions are evaluated in other two regions and an additional region completely from different MVS (Multi-view stereo) dataset; With the application to MVS 3D reconstruction, we concluded that the proposed SVM-based quality indicator outperforms simple pair selection strategies: 1) example-based method that simply ranks image pairs based on evaluated accuracy against known ground-truth data of a sub-region; 2) intersection-angle based method that randomly select pairs with intersection angles within 15-25°; 3) the method of (Facciolo et al., 2017) that pre-select pairs with intersection and maximal off-nadir angles respectively within 5-45°, and below 40°, and then rank the selected pairs based on their acquisition time interval (shorter the higher rank). In addition, we show that the proposed SVM-based quality indicator is potentially scalable and can serve as a general satellite stereo-pair quality indicator by using it on other MVS datasets.

Through regression analysis we show that intersection angles, sun-angle differences and matching-completeness (a post-matching parameter) are important factors for stereo pair quality: In general larger-intersection angles yield higher per-point accuracy; while for dense reconstruction for urban area, a relatively smaller intersection angle generally yields better and more complete DSM, it is therefore scene dependent; Sun-angle difference plays an equivalent and sometimes a more important role in stereo reconstruction, as it may cast different directional illumination to anisotropic/specular ground surfaces which can be challenging for stereo algorithms; matching-completeness can be algorithm dependent but in general can serve as a good indicator for stereo quality if pre-matching is possible in certain scenarios (e.g. computational time is tolerable).

The contribution of this paper is mainly two-fold:

1. We experimentally demonstrated the complex association between critical meta-parameters of satellite stereo pairs and their potentially achievable accuracy, and deliver important conclusions serving as key references for future researches on satellite stereo reconstruction.
2. We have proposed a general quality indicator based on a SVM model trained using our data and demonstrated the effectiveness of this model by applying to a MVS 3D reconstruction pipeline.

This proposed research has thoroughly investigated the stereo-pair dense matching quality using over 1,000 stereo pairs (data points) under varying acquisition parameters. Through the analyses we answer the question of how to select good satellite stereo pairs for 3D reconstruction, and additionally provide a means of using our data for stereo pair quality indicator using an SVM model. To further validate our analysis and the proposed matching quality indicator, we expect our future works include higher volume of data that encapsulates more terrain types for analysis and testing and improving our proposed quality indicator.

7. Acknowledgments

This work is partially supported by the Office of Naval Research (Award No. N000141712928). The dataset used for the analysis is created for the IARPA multi-view stereo 3D mapping Challenge and the Commercial satellite imagery in the MVS benchmark data set was provided courtesy of DigitalGlobe. The author would like to acknowledge Yilong Han and Xu Huang for assisting organizing parts of the figures of this manuscript.

8. References

Bay, H., T. Tuytelaars and L. Van Gool, 2006. Surf: Speeded up robust features. Computer vision–ECCV 2006 404-417.

Bosch, M., Z. Kurtz, S. Hagstrom and M. Brown, 2016a. A Multiple View Stereo Benchmark for Satellite Imagery. Proceedings of the IEEE Applied Imagery Pattern Recognition (AIPR) Workshop, October 2016,

Bosch, M., Z. Kurtz, S. Hagstrom and M. Brown, 2016b. A multiple view stereo benchmark for satellite imagery. In: Applied Imagery Pattern Recognition Workshop (AIPR), 2016 IEEE, pp. 1-9.

Bosch, M., A. Leichtman, D. Chilcott, H. Goldberg and M. Brown, 2017. Metric Evaluation Pipeline for 3d Modeling of Urban Scenes. The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 42 239.

Boyd, S. and L. Vandenberghe, 2004. Convex optimization. Cambridge university press,

Carl, S., S. Bärisch, F. Lang, P. D'angelo, H. Arefi and P. Reinartz, 2013. Operational generation of high resolution digital surface models from commercial tri-stereo satellite data. In: Photogrammetric Week, pp. 261-269.

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

d'Angelo, P., C. Rossi, C. Minet, M. Eineder, M. Flory and I. Niemeyer, 2014. High Resolution 3D Earth Observation Data Analysis for Safeguards Activities. In: Symposium on International Safeguards, pp. 1-8.

Dowman, I. and P. Michalis, 2003. Generic rigorous model for along track stereo satellite sensors. In: ISPRS Workshop on High Resolution Mapping from Space, pp.

Facciolo, G., C. De Franchis and E. Meinhardt-Llopis, 2017. Automatic 3D reconstruction from multi-date satellite images. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 57-66.

Fischler, M. A. and R. C. Bolles, 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM 24 (6), 381-395.

Foody, G. M. and A. Mathur, 2004. A relative evaluation of multiclass image classification by support vector machines. IEEE Transactions on Geoscience and Remote Sensing 42 (6), 1335-1343.

Fraser, C. S. and H. B. Hanley, 2003. Bias compensation in rational functions for IKONOS satellite imagery. Photogrammetric engineering and remote sensing 69 (1), 53-58.

Gehrke, S., K. Morin, M. Downey, N. Boehrer and T. Fuchs, 2010. Semi-global matching: An alternative to LIDAR for DSM generation. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Calgary, AB, 38 (B1) 6.

Geiger, A., P. Lenz and R. Urtasun, 2012. Are we ready for autonomous driving? the kitti vision benchmark suite. In: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, pp. 3354-3361.

Gruen, A. and D. Akca, 2005. Least squares 3D surface and curve matching. ISPRS Journal of Photogrammetry and Remote Sensing 59 (3), 151-174.

Hirschmüller, H., 2005. Accurate and efficient stereo processing by semi-global matching and mutual information. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 807-814.

Hirschmüller, H., 2008. Stereo processing by semiglobal matching and mutual information. IEEE Transactions on Pattern Analysis and Machine Intelligence 30 (2), 328-341.

Huber, P. J., 2011. Robust statistics. Springer, 1248-1251.

IARPA, 2016. IARPA Multi-view stereo 3D mapping challenge, https://www.iarpa.gov/challenges/3dchallenge.html. 2017 Feb 22

Jiang, B., X. Zhang and T. Cai, 2008. Estimating the confidence interval for prediction errors of support vector machine classifiers. Journal of Machine Learning Research 9 (Mar), 521-540.

Jordan, G., 2007. Adaptive smoothing of valleys in DEMs using TIN interpolation from ridgeline elevations: An application to morphotectonic aspect analysis. Computers & geosciences 33 (4), 573-585.

Kolmogorov, V. and R. Zabih, 2001. Computing visual correspondence with occlusions using graph cuts. In: Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on, pp. 508-515.

Krauß, T., P. D'Angelo and L. Wendt, 2018. Cross-track satellite stereo for 3D modelling of urban areas. European Journal of Remote Sensing 1-10.

Kuschk, G., P. d'Angelo, R. Qin, D. Poli, P. Reinartz and D. Cremers, 2014. DSM Accuracy Evaluation for the ISPRS Commission I Image matching Benchmark. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 40 195-200.

Lowe, D. G., 2004. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60 (2), 91-110.

Lu, G. Y. and D. W. Wong, 2008. An adaptive inverse-distance weighting spatial interpolation technique. Computers & geosciences 34 (9), 1044-1055.

Mathworks, 2018. Robust Regression, https://www.mathworks.com/help/stats/robustfit.html. Accessed July 2, 2018

Morgan, M., K. Kim, S. Jeong and A. Habib, 2004. Epipolar geometry of linear array scanners moving with constant velocity and constant attitude. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XXXV (B3), 1024-1029.

Pal, M. and P. Mather, 2005. Support vector machines for classification in remote sensing. International Journal of Remote Sensing 26 (5), 1007-1011.

Park, M.-G. and K.-J. Yoon, 2018. Learning and selecting confidence measures for robust stereo matching. IEEE Transactions on Pattern Analysis and Machine Intelligence

Pérez, P., M. Gangnet and A. Blake, 2003. Poisson image editing. ACM Transactions on graphics (TOG) 22 (3), 313-318.

Qin, R., 2017. Automated 3D recovery from very high resolution multi-view satellite images. In: ASPRS (IGTF) annual Conference, March 12-16, Baltimore, Maryland, USA, pp. 10.

Qin, R., 2014. Change detection on LOD 2 building models with very high resolution spaceborne stereo imagery. ISPRS Journal of Photogrammetry and Remote Sensing 96 (2014), 179-192.

Qin, R., 2016a. RPC Stereo Processor (RSP) –A software package for digital surface model and orthophoto generation from satellite stereo imagery. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences (to appear in ISPRS congress July 2016))

Qin, R., 2016b. RPC Stereo Processor (RSP) –A software package for digital surface model and orthophoto generation from satellite stereo imagery. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences III (1), 77-82.DOI: 10.5194/isprs-annals-III-1-77-2016

Qin, R., J. Tian and P. Reinartz, 2016. 3D change detection–Approaches and applications. ISPRS Journal of Photogrammetry and Remote Sensing 122 41-56.

Rothermel, M., K. Wenzel, D. Fritsch and N. Haala, 2012. SURE: Photogrammetric Surface Reconstruction from Imagery. In: Proceedings LC3D Workshop, Berlin, pp.

Scharstein, D. and R. Szeliski, 2014. Middlebury stereo vision page, http://vision.middlebury.edu/stereo/. (Accessed 03 December, 2014)

Scharstein, D. and R. Szeliski, 2002. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. International Journal of Computer Vision 47 (1-3), 7-42.

Seki, A. and M. Pollefeys, 2017. SGM-Nets: Semi-global matching with neural networks. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, pp. 21-26.

Smola, A. J. and B. Schölkopf, 2004. A tutorial on support vector regression. Statistics and computing 14 (3), 199-222.

SpaceNet, 2018. SpaceNet on Amazon Web Services (AWS). https://spacenetchallenge.github.io/datasets/datasetHomePage.html. Last modified April 30, 2018. Accessed Feb 5, 2019.

Street, J. O., R. J. Carroll and D. Ruppert, 1988. A note on computing robust regression estimates via iteratively reweighted least squares. The American Statistician 42 (2), 152-154.

Szeliski, R., 2010. Computer vision: algorithms and applications. Springer Science & Business Media,

Wang, L., 2005. Support Vector Machines: theory and applications. Springer, 431.

Wang, M., F. Hu and J. Li, 2011. Epipolar resampling of linear pushbroom satellite imagery by a new epipolarity model. ISPRS Journal of Photogrammetry and Remote Sensing 66 (3), 347-355.

Waser, L., E. Baltsavias, K. Ecker, H. Eisenbeiss, E. Feldmeyer-Christe, C. Ginzler, M. Küchler and L. Zhang, 2008. Assessing changes of forest area and shrub encroachment in a mire ecosystem using digital surface models and CIR aerial images. Remote Sensing of Environment 112 (5), 1956-1968.

Zabih, R. and J. Woodfill, 1994. Non-parametric local transforms for computing visual correspondence. In: Proceedings of European Conference on Computer Vision, Stockholm, Sweden, May, pp. 151-158.

Zbontar, J. and Y. LeCun, 2016. Stereo matching by training a convolutional neural network to compare image patches. Journal of Machine Learning Research 17 (1-32), 2.

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