

TEXTURE-AWARE DENSE IMAGE MATCHING USING TERNARY CENSUS TRANSFORM

Han Hu^{a,b,*}, Chongtai Chen^b, Bo Wu^a, Xiaoxia Yang^c, Qing Zhu^{b,d}, Yulin Ding^b

^aDepartment of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

^bFaculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Gaoxin West District, Chengdu, China

^cCollege of Earth Sciences, Chengdu University of Technology, Chengdu 610059, China

^dCollaborative Innovation Center for Geospatial Technology, 129 Luoyu Road, Wuhan, China

Commission III, WG III/1

KEY WORDS: Dense Image Matching, Texture aware, Census Transform, Local Ternary Pattern, SGM, Matching Cost

ABSTRACT:

Textureless and geometric discontinuities are major problems in state-of-the-art dense image matching methods, as they can cause visually significant noise and the loss of sharp features. Binary census transform is one of the best matching cost methods but in textureless areas, where the intensity values are similar, it suffers from small random noises. Global optimization for disparity computation is inherently sensitive to parameter tuning in complex urban scenes, and must compromise between smoothness and discontinuities. The aim of this study is to provide a method to overcome these issues in dense image matching, by extending the industry proven Semi-Global Matching through 1) developing a ternary census transform, which takes three outputs in a single order comparison and encodes the results in two bits rather than one, and also 2) by using texture-information to self-tune the parameters, which both preserves sharp edges and enforces smoothness when necessary. Experimental results using various datasets from different platforms have shown that the visual qualities of the triangulated point clouds in urban areas can be largely improved by these proposed methods.

1. INTRODUCTION

Recent innovations in aerial camera systems (Fritsch and Rothmel, 2013) and dense image matching (DIM) (Haala, 2013) approaches have led to the development of more image-based large-scale 3D city reconstruction. Modern DIM approaches are generally formulated as a discrete optimization problems on the disparity space image (Scharstein and Szeliski, 2002). As 2D optimization on the disparity image is inherently NP-complete, the computational complexities are prohibitively high for large-scale aerial applications, even with relatively efficient approximate optimization algorithms, such as loopy belief propagation (Sun et al., 2003) and graph cut (Kolmogorov and Zabih, 2001). Fortunately, it is possible to separately carry out one-dimensional optimization for each scan line through dynamic programming (DP) in polynomial time (Birchfield and Tomasi, 1999). The industry proven method of Semi-global Matching (SGM) (Hirschmuller, 2008) extends DP in multiple directions, and is effective in both airborne and satellite-borne applications (Alobeid et al., 2010; Remondino et al., 2014).

The quality of photogrammetric point clouds is widely acknowledged to be inferior to that of laser scanning (Gerke, 2009; Nex and Gerke, 2014), particularly regarding noise and the preservation of sharp features. Insufficient texture and geometric discontinuity are crucial problems in DIM algorithms, particularly in urban areas with artificial structures and sharp edges. First, the lack of texture causes ambiguities in the similarity measurement (referred to as matching cost) of the local appearances of image intensities (Yoon and Kweon, 2006). Second, unlike sparse feature matching, most DIM methods

explicitly integrate smoothness constraints to obtain dense point clouds in textureless areas. However, the smooth assumption is contradictory to the preservation of sharp features, which is crucial for its successful application. The balance between matching cost and smoothness is usually tuned through manually assigned parameters (Kolmogorov and Zabih, 2001; Sun et al., 2003; Hirschmuller, 2008), but in complex urban scenes, it is much more difficult to reach a compromise between noise and sharp feature preservation through a single set of parameters.

To address these problems, SGM is extended in this study, in terms of both the matching cost computation and disparity optimization. First, a ternary census transform is proposed to handle the ambiguities in textureless areas. Compared to the binary census transform (Birchfield and Tomasi, 1998), the proposed method also considers the equality of pixel intensities, as textureless areas typically have similar grayscale values and the relative order is more easily affected by random noise. Second, a texture-aware SGM is proposed to adaptively tune the parameters, to both enforce smoothness and preserve sharp edges. Because textureless areas not only represent similar intensities but also disparity continuities and vice versa (Stentoumis et al., 2014; Yang, 2015), this study quantitatively measures the texture information then self-tunes the parameters of SGM accordingly.

2. RELATED WORKS

With recent developments in aerial camera systems (Fritsch and Rothmel, 2013) and the automation of 3D reconstruction, DIM have now been a hot topic in the photogrammetry

* Corresponding Author. Email: lshuhan@polyu.edu.hk

community (Cavegn et al., 2014; Kuschka et al., 2014; Krombach et al., 2015). Most DIM methods typically have four stages (Scharstein and Szeliski, 2002; Yang, 2015): 1) matching cost computation, 2) matching cost aggregation, 3) disparity computation/optimization and 4) disparity refinement. Taking traditional area-based matching as an example, the matching cost is the square difference of pixel intensities, matching cost aggregation is generally accomplished using the Normalized Correlation Coefficient (NCC) in a rectangular window, the disparity computation is carried out through a local winner-take-all (WTA) operation in a small search window and screening with a simple threshold, and the final step typically involves sub-pixel interpolation and other post processing. DIM methods are classified as either local or global, distinguished by the use of smoothness constraints and the inference of disparities (Scharstein and Szeliski, 2002; Lazaros et al., 2008). The former depends only on local intensity values, while global reasoning is generally used in the latter for disparity optimization.

DIM methods in both approaches generally face common problems, such as insufficient texture and the preservation of sharp features, which often appear in complex urban areas (Remondino et al., 2014). Textureless areas will cause ambiguities in determining matching cost and therefore magnify errors from the random noise of pixel intensities. Because of the difficulties to obtain local optimum in matching of textureless areas, DIM methods will generally fail in this area without smoothness constraint (Lazaros et al., 2008). The simplest strategy to exploit smoothness is to aggregate matching cost in a larger window (Xin et al., 2012), as the aggregation window implicitly uses the assumption of the same disparity value (Yang, 2015). It is also possible to enforce smoothness through segmentation: the images are first clustered into homogenous segments using color information, and pixels in each segment are implicitly assumed to share similar or smoothed disparities (Bleyer and Gelautz, 2004; Hong and Chen, 2004; Zhang et al., 2009). Similar effects can also be achieved through plane fitting, where the plane proposals can originate from color consistencies (Klaus et al., 2006) or initial feature matches (Wu et al., 2011; Wu et al., 2012; Sinha et al., 2014). In the global scheme, the smoothness constraints can be explicitly enforced by an energy term that penalizes neighbor pixels with different disparities (Kolmogorov and Zabih, 2001; Sun et al., 2003; Hirschmuller, 2008).

Insufficient texture requires the DIM algorithms to enforce smoothness and use global reasoning to alleviate noise and the ambiguities of matching cost. However, a smoothness constraint also flattens sharp features or causes the loss of boundary points due to failures in cross-checking (Remondino et al., 2014). Small rather than large aggregation windows are preferred for sharp features, or in extreme cases, pixelwise cost without aggregation can be used (Scharstein and Szeliski, 2002), such as absolute difference, sampling-insensitive absolute difference (Birchfield and Tomasi, 1998), mutual information (Hirschmuller, 2008), and census transform (Zabih and Woodfill, 1994). Weighing the matching window to avoid edges is another widely used strategy. For example, Yoon and Kweon (2006) assign appropriate weight by color similarity and spatial proximity, which transpires to be analogous to the bilateral filter (Hirschmüller and Scharstein, 2009). Bilateral filtering can also be applied to the minimum spanning tree, to achieve non-local aggregation and linear time complexity (Yang,

2015). Weighting can also be applied to non-parametric census transform through duplicating code (Spangenberg et al., 2013) or adaptively modifying the aggregation window (Stentoumis et al., 2014). It is possible to preserve geometric discontinuities through global methods by loosening the smoothness constraints through parameter tuning, e.g., P_1 and P_2 for SGM (Hirschmuller, 2008), and λ for graph cut (Kolmogorov and Zabih, 2001).

A textureless area will inevitably decrease the performance of the matching cost *per se*. The most intuitive solution is to increase the resiliency of matching cost to noise and other conditions. In a comprehensive comparison of many matching costs, the census transform has been claimed to demonstrate the best and most robust overall performance (Hirschmüller and Scharstein, 2009). However, as relative intensity order does not take equality into account in computing the binary census transform, in a textureless area with many similar pixel intensities, a high level of incorrect ordering through random noises, camera qualities, and vignetting are expected. Using ternary order, which accounts for equality, So in this paper, we extend the binary order results with ternary order that takes equality into account (Tan and Triggs, 2010; Murala et al., 2012), and modify the pattern accordingly to allow for efficient computation.

Using adaptive filtering to preserve sharp features is in general only suitable for local methods, which are usually less accurate than global methods (Stentoumis et al., 2014). However, in global methods it is contradictory to both enforce smoothness and preserve discontinuities in complex scenes by a single parameter set. Following the suggestion to adaptively tune parameters in complex scenes (Remondino et al., 2014), we incorporate texture information to achieve both smoothness and discontinuity in different scenarios.

3. METHODS

3.1 Overview

Because SGM is an industry proven DIM algorithm that demonstrates superior precision and production-level efficiency (Alobeid et al., 2010; Remondino et al., 2014), the method proposed in this paper is based on SGM. As the basic concept is compatible with any global method using pixelwise matching cost, similar strategies can be easily applied to other DIM algorithms. The overall workflow of the proposed method is summarized in Figure 1. The method differs from others as it uses ternary census transform for matching cost computation and incorporates texture information into SGM to self-tune the parameters for the disparity computation.

DIM methods typically use an epipolar relationship to constrain the correspondence search in a one-dimensional direction. The epipolar relationship can be calculated dynamically in disparity optimization, but for efficient implementation we pre-rectify the stereo image pairs to obtain horizontal epipolar images. For frame images, a homography transform is used for the rectification (Fusiello et al., 2000). For pushbroom satellite images, though the epipolar geometry is a quadratic curve instead of a line, in a relative small region, a simple affine transform can also be used (Wang et al., 2010), due to the stability of the satellite platform during image acquisition.

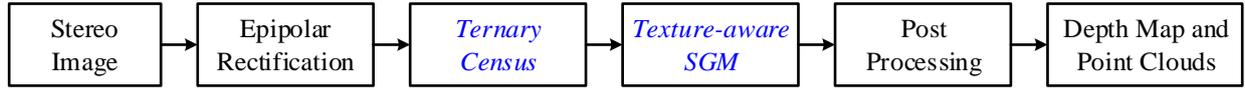


Figure 1. Workflow of the proposed texture-aware SGM using a ternary census transform.

After rectification we convert the epipolar image into corresponding codes using the proposed ternary census transform, where each pixel encodes a 64-bit census code (detailed in Section 3.2). The matching cost is measured using the hamming distance and dynamically computed in a single instruction¹, rather than reallocated in the cost volume. Forward and backward texture-aware SGM are then conducted to achieve the final disparity image in a coarse-to-fine scheme (detailed in Section 3.3). SGM can be formulated as DP in multiple directions, and in a single direction it solves the following problem:

$$L_r(p, d) = C(p, d) + \min(L_r(p-r, d), L_r(p-r, d-1) + P_1, L_r(p-r, d+1) + P_1, \min_i L_r(p-r, i) + P_2) - \min_k L_r(p-r, k) \quad (1)$$

where P_1 and P_2 control the penalties for adjacent pixels with different disparities and other parameters are found in (Hirschmüller, 2008). P_1 is imposed on small and smooth changes of disparity and P_2 on larger and acute changes. In this study we adaptively tune these parameters according to the texture information, to achieve less noise in textureless areas and to preserve geometric discontinuities in shared edges.

Potential outliers in the pairwise disparity map are first removed using speckle filters (Hirschmüller, 2008). The disparity map is then refined by parabola interpolation, using the matching cost of three adjacent pixels (Wang, 1990). The invalid pixels that fail the cross-check are left unchanged without interpolation, as suggested by the original SGM used for aerial images (Hirschmüller, 2008), as this may cause further outliers that are smooth and difficult to remove.

For multiple overlapping views of the same area, multiple disparity maps are fused into the depth map defined in the original image space. The fusion is carried out using a weighted average, and weights for every pair are determined by the length of the baseline. When converting the depth maps to corresponding point clouds, gridded median filters are used. This process also resamples the point clouds and removes potential duplicate points.

3.2 Ternary Census Transform

The census transform is a non-parametric matching cost, which relies only on the ordering of pixel intensities and is therefore invariant to any radiometric changes that preserve the intensity order (Hirschmüller and Scharstein, 2009). For any given rectangular window, a pixel will be encoded into a bit string, which concatenates the relative order of neighboring pixels, as described in Figure 2a. The census transform is defined with

$$CT(x, y) = \bigotimes_{i=-n' \atop i=n'}^{i=n' \atop j=m'} \bigotimes_{j=-m' \atop j=m'} b(I(x, y), I(x+i, y+j)) \quad (2)$$

where \bigotimes is the concatenation operator, $CT(\bullet)$ is the code image after transformation, $I(\bullet)$ is the epipolar image, and $b(\bullet)$ is the binary operator as in the following.

$$b(x, y) := \begin{cases} 0 & \text{if } x > y \\ 1 & \text{if } x \leq y \end{cases} \quad (3)$$

As previously described, small variations in textureless areas with similar pixel intensities will cause differences in the binary code. The sources of radiometric variations may be illumination settings, image noise, vignetting, or even jpeg compression. To enhance the robustness of the matching cost in textureless areas, we take the equality of the intensity comparison into account. As Figure 2b shows, instead of using the binary operator in Equations (2) and (3), the proposed census transform takes the ternary operator, following the local ternary pattern concept (Tan and Triggs, 2010; Murala et al., 2012),

$$CT(x, y) = \bigotimes_{i=-n' \atop i=n'}^{i=n' \atop j=m'} \bigotimes_{j=-m' \atop j=m'} t(I(x, y), I(x+i, y+j)) \quad (4)$$

$$t(x, y, \delta) := \begin{cases} 00 & \text{if } x > y + \delta \\ 01 & \text{if } y - \delta \leq x \leq y + \delta \\ 11 & \text{if } x < y - \delta \end{cases}$$

where δ is the threshold for equality.

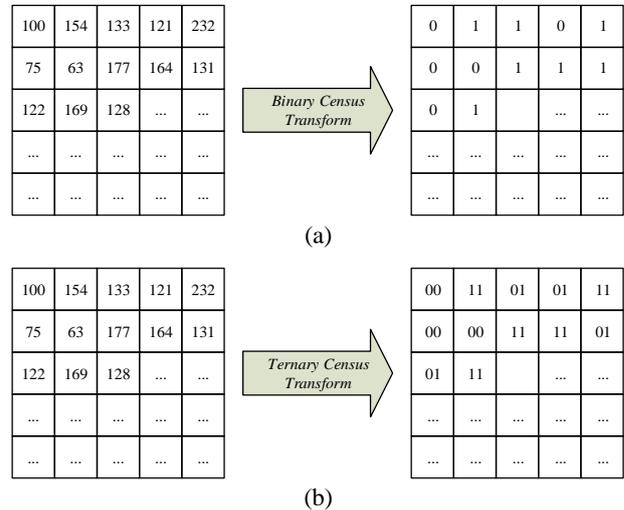


Figure 2. Illustration of census transform on a 5×5 grid; (a) Binary census transform and (b) ternary census transform.

A single comparison results in two bits rather than one, so it appears that the storage and computation may be two-fold in the binary census transform. However, unlike the traditional pixel-by-pixel comparison carried out with a 9×7 grid (Hirschmüller and Scharstein, 2009), we use a predefined pattern on a 9×9 grid, where only the shaded squares are evaluated, as shown in Figure 3. This sparse and symmetrical pattern has two possible advantages: 1) a dense grid may have a close correlation between adjacent pixels, and 2) the pattern is designed to follow the Gaussian kernel, thus giving more weight to pixels close to the center. As only 32 grids are used for the evaluation, the total length for a single pixel is 64-bit and the matching cost can be evaluated in a single instruction.

¹ The *popcnt* instruction in SSE4.2 is able to calculate the non-zero bit in an unsigned 64-bit integer.

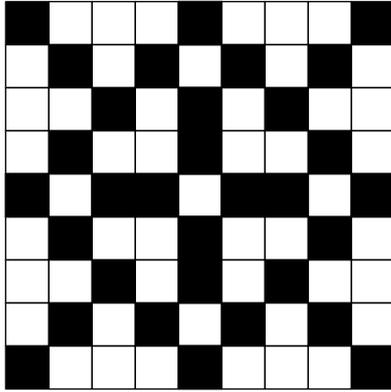


Figure 3 A symmetrical pattern on a 9×9 grid. For a single pixel, the proposed ternary census transform results in 64-bit binary codes.

It should be noted that the matching cost of the ternary census transform is also measured by the Hamming distance, which is the number of different bits in the coded string. The matching cost between 00 and 11 is 2, and between equality and inequality is 1. The ternary census transform will therefore result in an intermediate cost for smaller differences.

3.3 Texture-Aware Semiglobal Matching

Texture is the frequency of color variation and represents the level of image detail (Malik et al., 2001). The premise of texture-aware SGM is that color variation is in direct proportion to the variation of the disparity and thus the smoothness. The richer the texture, the larger the parameter P in Equation (1) should be. The texture information can be quantitatively measured by information entropy (Zhu et al., 2007), a number of filters with different orientations and scales (Malik et al., 2001), and by simple gradient analysis (Xue et al., 2014). Textureless areas are known to result in a low gradient and variation in color appearance, so this study uses the above two cues to quantitatively gauge the texture information, as;

$$g(x, y) = \frac{1}{2N} \sum_{(u,v)}^{W(x,y)} |I(u, v) - I(u+1, v)| + |I(u, v) - I(u, v+1)|$$

$$\sigma(x, y) = \sqrt{\frac{1}{N} \sum_{(u,v)}^{W(x,y)} (I(u, v) - \bar{I}(x, y))^2} \quad (5)$$

$$T(x, y) = w_1 g(x, y) + w_2 \sigma(x, y)$$

where $g(x, y)$ is the average of gradient in both directions and $\sigma(x, y)$ is the standard deviation of intensities. $W(x, y)$ is a local window and in this study the same 9×9 grid is used. Finally, $T(x, y)$ is the weighted texture information using both of these factors, where $w_1=w_2=0.5$ is used in this paper.

The original SGM exploits this idea by simple adaption with derivatives (Hirschmuller, 2008) as $P_2=P_2'/|I_p-I_q|$ and another excellent SGM implementation, SURE, uses canny edges to adaptively change between two values as P_2 or P_2-P_3 (Rothermel et al., 2012). However, in real applications we have found that the derivatives are quite sensitive to noise, and the

canny edges require two parameters to explicitly binarize the image. A comparison of the above two approaches and the proposed method is shown in Figure 4. It should be noted that the derivatives are essentially sensitive to noise, and using the same thresholds, the canny method sometimes omits edges or results in dashed edges, illustrated by the red rectangles. However, it may also cause noise edges, as shown by the yellow rectangles.

After obtaining the texture information of the epipolar images, an intuitive solution for adaptive parameter selection is linear interpolation between the user-defined minimum and maximum. However, this requires a further temporary buffer to save the parameters, and an extra load operation for Single Instruction Multiple Data (SIMD) optimization. Similar to SURE (Rothermel et al., 2012), we therefore also binarize the texture and thus the parameter. Manual parameter-tuning is a time-consuming task, so an adaptive thresholding strategy using local Gaussian weighted mean is used. An inconsistency of even one pixel in disparity has been found to sometimes cause more than 0.2 meters of perturbations in the point clouds, so we also adapt P_1 in Equation (1) rather than only P_2 , as in previous studies (Hirschmuller, 2008; Rothermel et al., 2012). We can then obtain a parameter mapping function for each pixel $P_1(p)$ and $P_2(p)$, which are linear mapping between user defined ranges as detailed in experiments.

4. EXPERIMENTAL RESULTS

4.1 Test data description

To evaluate performance and efficiencies, we tested the proposed methods on three different datasets: Jinyang, Zurich (Cavegn et al., 2014), and Vaihingen (Cramer, 2010), all of which cover built-up urban areas. The images in Vaihingen are collected by large-frame aerial cameras with a relatively large baseline-height-ratio with nadir views. The radiometric qualities of the images can also be superior in this dataset. The pan-sharpening colour infrared images are used for matching.

The other two datasets both consist of aerial oblique images, but the configuration and tilt angles are quite different. In the Jinyang dataset, the tilt angle is 45° and the focal length is about 50 mm and 80 mm for nadir and oblique views, respectively. For the Zurich dataset, the tilt angle is 35° and the focal length is 50 mm for all the cameras. Both datasets have RGB color images, but the radiometric resolutions are different: 8-bit for the Jinyang dataset and 16-bit for the Zurich dataset. A brief overview of the test data is summarized in Table 1.

Dataset	Sensor	Focal(mm)	GSD(cm)	Type
Vaihingen	DMC	120	8	Nadir
Jinyang	SWDC-5	50/80	8 (8-16)	Oblique
Zurich	RCD-30	50/50	6 (6-13)	Oblique

Table 1. Description of the test data.

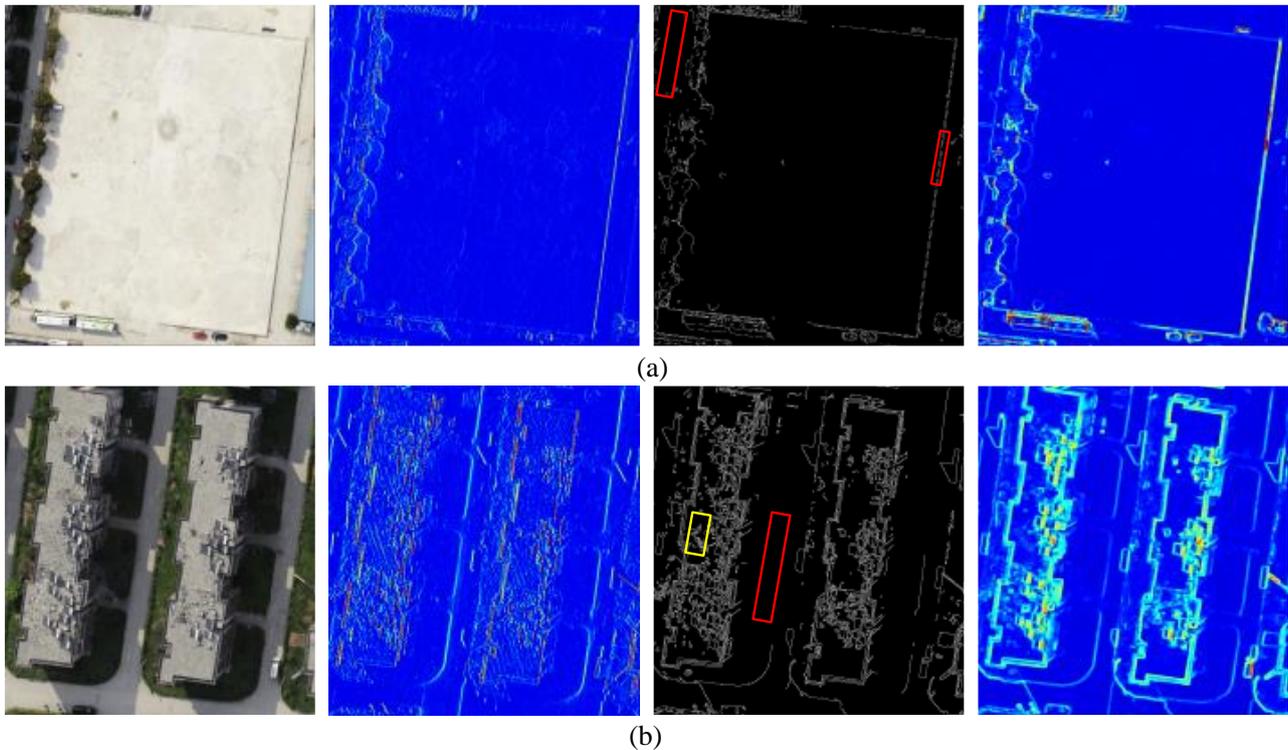


Figure 4. Comparison of different texture representations. From left to right: original images, derivatives implemented by the original SGM (Hirschmuller, 2008), canny edges implemented in SURE (Rothermel et al., 2012), and the proposed texture representation.

Top row: comparison in a cement square; Bottom row: residential building roof with small structures.

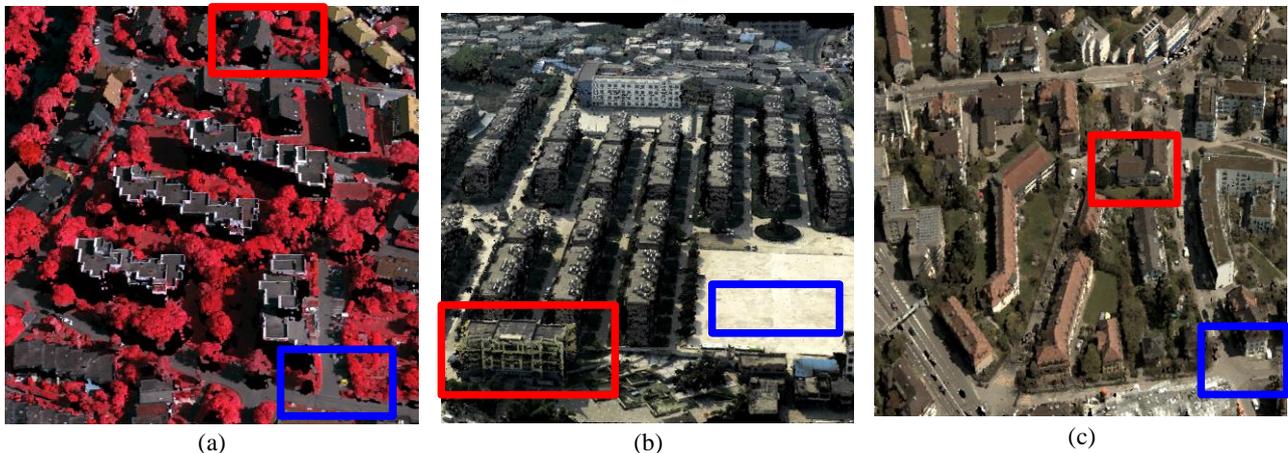


Figure 5. Overview of the point clouds in the evaluated areas. (a) Vaihingen dataset, (b) Jinyang dataset, and (c) Zurich dataset. The red and blue rectangles show the evaluated building and road areas, respectively.

4.2 Results

To analyze the effectiveness of the proposed methods on textureless areas and geometric discontinuities, we evaluate the matching results on the above four datasets using three parameter configurations; smaller parameter, larger parameters, and adaptive parameters. We then compare the matching results on typical scenarios, to visually evaluate the amount of noise and the quality of point clouds on the sharp edges. In all the experiments, no matter what radiometric resolutions and camera types, the matching parameters are all set the same at [28, 56] and [70, 140] for P_1 and P_2 , respectively. Figure 5 shows the overview of the point clouds in the evaluated area and highlights the textureless and geometric discontinuities with rectangles.

Figure 6 through Figure 8 show the matching results for the four different datasets. In each figure, the first column shows the subset of the images. The three right-hand columns (b), (c), and (d) show detailed results for smaller, larger, and adaptive parameters, respectively. In each column, the top row shows the results of a relatively smooth area from the blue rectangle, and the bottom row shows the results of a building area.

Figure 6 shows the matching results from the two large-frame cameras, which capture the nadir images. In the top row of the figure detailed matching results for smooth road areas with Delaunay triangulation are shown. As expected, smaller parameters will generate significantly more noise in the planar roads than larger parameters, found when results in columns (b) and (c) are compared. The noise may cause problems in two areas: 1) random noise in image intensities, which leads to an

incorrect matching cost; 2) a non-optimal matching cost, which causes small variations in sub-pixel interpolation. The horizontal plane in aerial nadir images will generate almost no disparity change, so it is reasonable to penalize both P_1 and P_2 . The adaptive parameter cases shown in column (d) share the same qualities as discussed above, as roads are generally labeled as textureless, even with road markers. Another interesting

effect in textureless areas is the stripped effect shown in Figure 6, which also appears in the SGM implementation of SURE. This effect is rooted in the nature of DP itself, except that the strip may have different orientations, rather than only a horizontal direction.

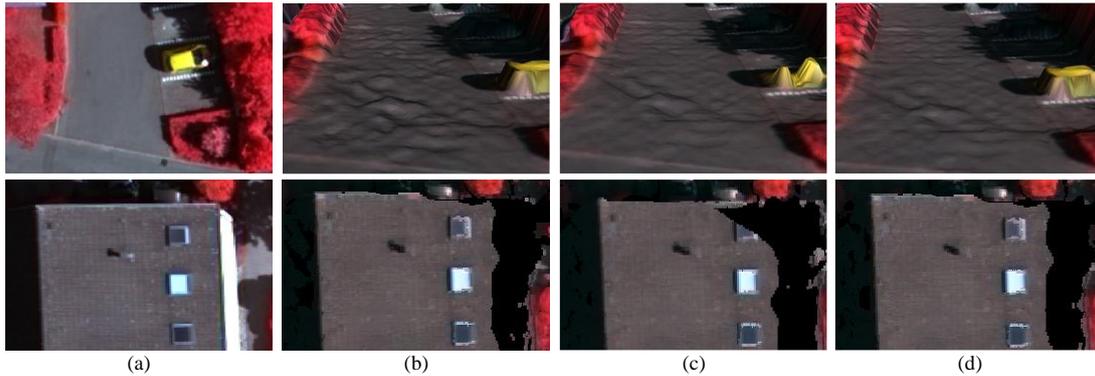


Figure 6. Matching results of Vaihingen dataset. a) enlarged images, b) results of small parameters ($P_1=28$ and $P_2=70$), c) results of larger parameters ($P_1=56$ and $P_2=140$), and d) results of the adaptive parameter selection. The top row shows textureless areas from the blue rectangles in Figure 5 and the bottom row shows geometric discontinuities from the red rectangles.

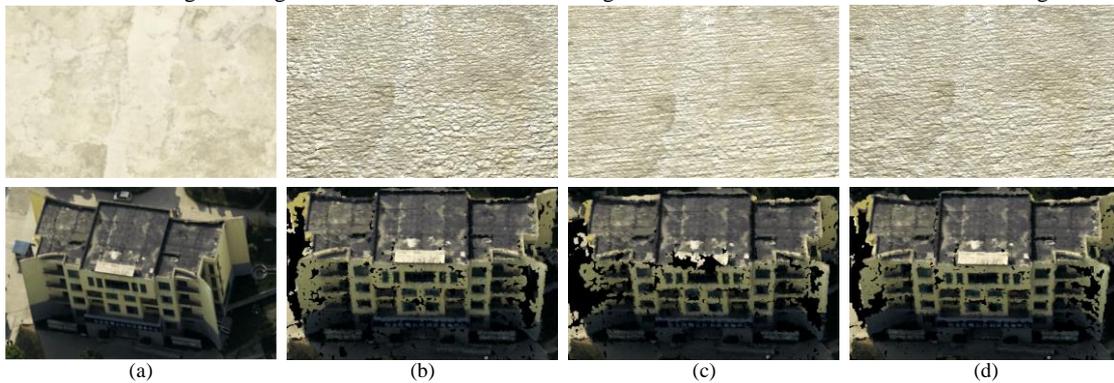


Figure 7. Matching results of Jinyang dataset. The symbols are the same as Figure 6.

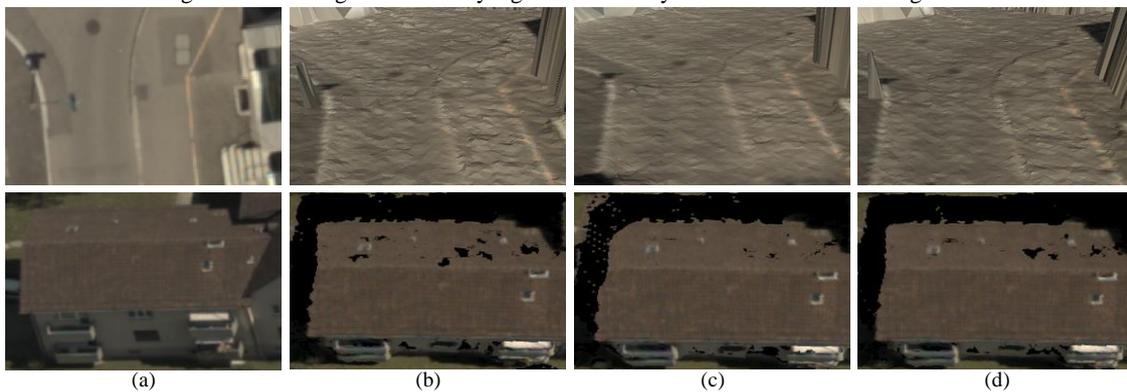


Figure 8. Matching results of Zurich dataset. The symbols are the same as Figure 6.

The bottom row of Figure 6 shows detailed point clouds over a building area. Larger parameters are found to suffer more from losses of matched points than smaller parameters, particularly in the corner area, as the disparity here is not consistent in all SGM scan directions. By over-penalizing for inconsistent disparities, the final matching results will deviate from the matching cost and create a smoothed disparity map; these disparities will probably fail the cross-check and be removed from the final results. By using smaller parameters, the discontinuities can be better preserved, and as the texture variation between roof and ground are much more significant,

they can also be used to derive the adaptive results in column (d).

The oblique cases provide similar results, as shown in Figure 7 and Figure 8. Here, the horizontal plane is actually slanted in the camera view, so the disparity will continuously change. To be consistent, we also adaptively tune parameter P_1 in this implementation, and the results prove that this strategy will generate better and smoother results in the horizontal planar square and for roads, as shown in the top row.

Airborne LiDAR data is also available in the Vaihingen dataset, so profile analyses are also presented, as shown in Figure 9. The densities of the two datasets are not the same, with an average interval between adjacent points for the DIM point clouds of about $0.1m$, and of about $0.3m$ for LiDAR. When sampling the profiles, we use the larger space as the LiDAR dataset. The profile of the LiDAR data is smoother in the planar area than the photogrammetric point clouds, as shown in Figure 9a, and also preserves better and sharper edges, shown in Figure 9b. For smaller parameters, the variation in the planar area is larger than that of the proposed texture-aware SGM, as the right-hand side of Figure 9a shows. For larger parameters, part of the edge of the building is lost and some sharp features are severely smoothed, as matches on the edges of buildings will not pass the cross-check. Furthermore, because the point clouds from SGM are coarsely sampled, the noises effect are smoothed for the profile analyses.

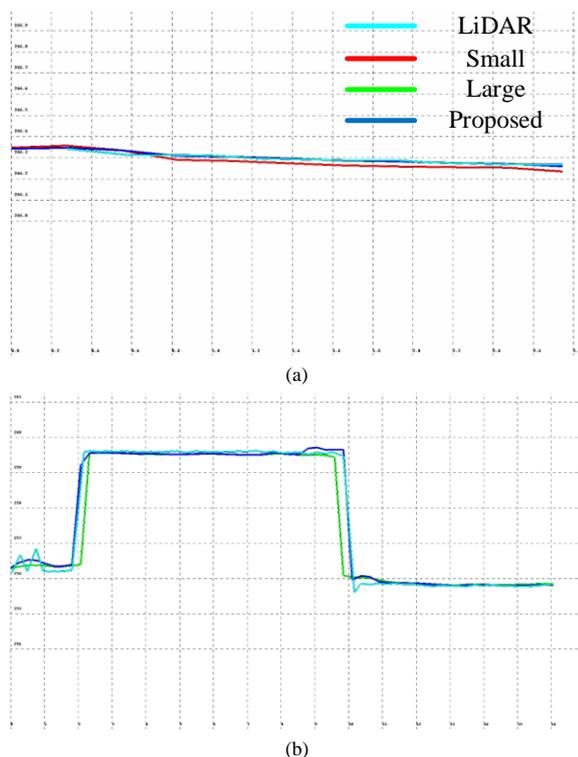


Figure 9. Comparisons of profile analyses of the photogrammetric and airborne LiDAR point clouds in the Vaihingen dataset, on (a) a road area and (b) a building area.

5. CONCLUSION

The aim of this study is to improve matching quality in textureless areas and at geometric discontinuities, using a texture-aware DIM method, with a ternary census transform for pixelwise cost computation. This method is based on SGM, but the concept can be generalized and easily adapted to other matching channels. The ternary census transform is designed to handle insufficient texture, where pixel intensities are similar and prone to small noises, while considering equality in the intensity comparison and encoding the results in two bits. Texture information is quantitatively measured and inserted into the cost aggregation step of SGM, and adaptively tunes the SGM parameters P_1 and P_2 . Finally, various camera systems and configurations are used in our extensive experimental evaluation of three different datasets, which visually demonstrates that the proposed method improves the matching

quality, as to some extent the noise in textureless areas is reduced and more sharp features are preserved than in traditional SGM.

ACKNOWLEDGEMENT

This study was supported by the National Natural Science Foundation of China (No. 41471320 and 41501421). And the Vaihingen dataset was provided by the German Society for Photogrammetry, Remote Sensing and Geoinformation (DGPF) [Cramer, 2010]: <http://www.ifp.uni-stuttgart.de/dgpf/DKEP-Allg.html>.

REFERENCES:

- Alobeid, A., Jacobsen, K., Heipke, C., 2010. Comparison of matching algorithms for DSM generation in urban areas from Ikonos imagery. *Photogrammetric Engineering and Remote Sensing*, 76 (9), pp. 1041-1050.
- Birchfield, S., Tomasi, C., 1998. A pixel dissimilarity measure that is insensitive to image sampling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20 (4), pp. 401-406.
- Birchfield, S., Tomasi, C., 1999. Depth discontinuities by pixel-to-pixel stereo. *International Journal of Computer Vision*, 35 (3), pp. 269-293.
- Bleyer, M., Gelautz, M., 2004. A layered stereo algorithm using image segmentation and global visibility constraints. In: *International Conference on Image Processing (ICIP2004)*, IEEE, Singapore, pp. 2997-3000.
- Cavegn, S., Haala, N., Nebiker, S., Rothermel, M., Tutzauer, P., 2014. Benchmarking high density image matching for oblique airborne imagery. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Zurich, Switzerland, Vol. XL-3, pp. 45-52.
- Cramer, M., 2010. The DGPF-test on digital airborne camera evaluation overview and test design. *Photogrammetrie-Fernerkundung-Geoinformation*, 2010 (2), pp. 73-82.
- Fritsch, D., Rothermel, M., 2013. Oblique image data processing - potential, experience and recommendation. In: *Photogrammetric Week 2013*, Stuttgart, Germany, pp. 73-88.
- Fusiello, A., Trucco, E., Verri, A., 2000. A compact algorithm for rectification of stereo pairs. *Machine Vision and Applications*, 12 (1), pp. 16-22.
- Gerke, M., 2009. Dense matching in high resolution oblique airborne images. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Paris, France, Vol. XXXVIII, Part 3/W4, pp. 77-82.
- Haala, N., 2013. The landscape of dense image matching algorithms. In: *Photogrammetry Week*, Stuttgart, Germany, pp. 271-284.
- Hirschmuller, H., 2008. Stereo processing by semiglobal matching and mutual information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30 (2), pp. 328-341.
- Hirschmüller, H., Scharstein, D., 2009. Evaluation of stereo matching costs on images with radiometric differences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31

(9), pp. 1582-1599.

Hong, L., Chen, G., 2004. Segment-based stereo matching using graph cuts. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR2004)*, IEEE, Washinton, DC, USA, pp. 74-81.

Klaus, A., Sormann, M., Karner, K., 2006. Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure. In: *18th International Conference on Pattern Recognition (ICPR2006)*, IEEE, Hong Kong, pp. 15-18.

Kolmogorov, V., Zabih, R., 2001. Computing visual correspondence with occlusions using graph cuts. In: *9th IEEE International Conference on Computer Vision (ICCV2001)*, IEEE, Vancouver, British Columbia, Canada, pp. 508-515.

Krombach, N., Droschel, D., Behnke, S., 2015. Evaluation of stereo algorithms for obstacle detection with fisheye lenses. In: *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Toronto, Canada, Vol II, Part 1/W1, pp. 33-40.

Kuschka, G., D'Angelo, P., Qin, R., Polic, D., Reinartz, P., Cremers, D., 2014. DSM Accuracy Evaluation for the ISPRS Commission I Image Matching Benchmark. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Denver, Colorado, Vol. XL, Part 1, pp. 195-200.

Lazaros, N., Sirakoulis, G.C., Gasteratos, A., 2008. Review of stereo vision algorithms: from software to hardware. *International Journal of Optomechatronics*, 2 (4), pp. 435-462.

Malik, J., Belongie, S., Leung, T., Shi, J., 2001. Contour and Texture Analysis for Image Segmentation. *International Journal of Computer Vision*, 43 (1), pp. 7-27.

Murala, S., Maheshwari, R.P., Balasubramanian, R., 2012. Local Tetra Patterns: A new feature descriptor for content-based image retrieval. *IEEE Transactions on Image Processing*, 21 (5), pp. 2874-2886.

Nex, F., Gerke, M., 2014. Photogrammetric DSM denoising. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Zurich, Switzerland, Vol. XL, Part 3, pp. 231-238.

Remondino, F., Spera, M.G., Nocerino, E., Menna, F., Nex, F., 2014. State of the art in high density image matching. *The Photogrammetric Record*, 29 (146), pp. 144-166.

Rothermel, M., Wenzel, K., Fritsch, D., Haala, N., 2012. Sure: Photogrammetric surface reconstruction from imagery. In: *Proceedings LC3D Workshop, Berlin*, Berlin, Germany, pp. 1-9.

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

Sinha, S.N., Scharstein, D., Szeliski, R., 2014. Efficient High-Resolution Stereo Matching using Local Plane Sweeps. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR2014)*, IEEE, Columbus, Ohio, USA, pp. 1582-1589.

Spangenberg, R., Langner, T., Rojas, R., 2013. Weighted Semi-Global Matching and center-symmetric Census Transform for robust driver assistance. In: *15th International Conference on*

Computer Analysis of Images and Patterns, Springer Berlin Heidelberg, York, UK, pp. 34-41.

Stentoumis, C., Grammatikopoulos, L., Kalisperakis, I., Karras, G., 2014. On accurate dense stereo-matching using a local adaptive multi-cost approach. *ISPRS Journal of Photogrammetry and Remote Sensing*, 91, pp. 29-49.

Sun, J., Zheng, N., Shum, H., 2003. Stereo matching using belief propagation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25 (7), pp. 787-800.

Tan, X., Triggs, B., 2010. Enhanced local texture feature sets for face recognition under difficult lighting conditions. *IEEE Transactions on Image Processing*, 19 (6), pp. 1635-1650.

Wang, M., Hu, F., Li, J., 2010. Epipolar arrangement of satellite imagery by projection trajectory simplification. *The Photogrammetric Record*, 25 (132), pp. 422-436.

Wang, Z., 1990. Principle of Photogrammetry: with Remote Sensing. Press of Wuhan Technical University of Surveying and Mapping, Wuhan, p. 455.

Wu, B., Zhang, Y., Zhu, Q., 2011. A triangulation-based hierarchical image matching method for wide-baseline images. *Photogrammetric Engineering and Remote Sensing*, 77 (7), pp. 695-708.

Wu, B., Zhang, Y., Zhu, Q., 2012. Integrated point and edge matching on poor textural images constrained by self-adaptive triangulations. *ISPRS Journal of Photogrammetry and Remote Sensing*, 68, pp. 40 - 55.

Xin, L., Fangjie, Y., Honghong, Z., Xiufang, L., Dalei, S., Bingwei, W., 2012. Illumination-robust area-based stereo matching with improved census transform. In: *International Conference on Measurement, Information and Control (MIC2012)*, Harbin, China, pp. 194-197.

Xue, W., Zhang, L., Mou, X., Bovik, A.C., 2014. Gradient magnitude similarity deviation: a highly efficient perceptual image quality index. *IEEE Transactions on Image Processing*, 23 (2), pp. 684-695.

Yang, Q., 2015. Stereo Matching Using Tree Filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37 (4), pp. 834-846.

Yoon, K., Kweon, I.S., 2006. Adaptive support-weight approach for correspondence search. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28 (4), pp. 650-656.

Zabih, R., Woodfill, J., 1994. Non-parametric local transforms for computing visual correspondence. In: *3rd European Conference on Computer Vision (ECCV1994)*, Springer, Stockholm, Sweden, pp. 151-158.

Zhang, G., Jia, J., Wong, T., Bao, H., 2009. Consistent depth maps recovery from a video sequence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31 (6), pp. 974-988.

Zhu, Q., Wu, B., Wan, N., 2007. A filtering strategy for interest point detecting to improve repeatability and information content. *Photogrammetric Engineering and Remote Sensing*, 73 (5), pp. 547-553.