DENSE IMAGE MATCHING WITH TWO STEPS OF EXPANSION

Zuxun Zhang [a], Jia'nan He [a], Shan Huang [a], Yansong Duan [a, *]

[a] School of Remote Sensing and Information Engineering, Wuhan University, No.129 Luoyu Road, Wuhan, China - zhangzx@cae.cn, (hjn2010, huangsh)@whu.edu.cn

Commission III, WG III/1

KEY WORDS: Dense Image Matching, Two Frames, Expansion, Seed-and-Grow, No Geometric Constraint, Adaptive

ABSTRACT:

Dense image matching is a basic and key point of photogrammetry and computer version. In this paper, we provide a method derived from the seed-and-grow method, whose basic procedure consists of the following: First, the seed and feature points are extracted, after which the feature points around every seed point are found in the first step of expansion. The corresponding information on these feature points needs to be determined. This is followed by the second step of expansion, in which the seed points around the feature point are found and used to estimate the possible matching patch. Finally, the matching results are refined through the traditional correlation-based method. Our proposed method operates on two frames without geometric constraints, specifically, epipolar constraints. It (1) can smoothly operate on frame, line array, natural scene, and even synthetic aperture radar (SAR) images and (2) at the same time guarantees computing efficiency as a result of the seed-and-grow concept and the computational efficiency of the correlation-based method.

1. INTRODUCTION

Dense image matching is an important step of 3D reconstruction and many other applications, with which the depth of or corresponding information on a pixel is determined. It has been one of the research hotspots in both photogrammetry and computer version for a long time. Dense image matching is challenging because of occlusions, blurred boundaries of objects, low or repetitive textures, and illumination changes. Moreover, the calculation speed is another obstacle, either because of the number of images or the size of images or because of the requirements of real-time applications.

The goal of dense image matching is to rebuild a 3D scene model from 2D images. To achieve this goal, the corresponding relationship of two or more images captured from different viewpoints of a scene must be rebuilt. The key and inherent problem of dense image matching is the extraction of corresponding features or points from different images. To the best of our knowledge, the problem of mismatching in the area of low-texture and disparity discontinuities is far from being solved. Reasonable prior constraints are used in matching algorithms to obtain better match results.

The existing algorithms can be divided into local and global methods, according to Scharstein and Szeliski (Scharstein and Szeliski, 2002). Local methods are based on the disparity consistency constraint, which requires determining a suitable window shape, window size, and weight of each pixel in the window. Image-segmentation information is also used as prior knowledge in some methods, but the matching results of those approaches heavily depend on the segmentation results of an image. To reduce the parallax search range, local methods frequently adopt various constraints, which results in a strong dependence on prior knowledge. At the same time, they are sensitive to the ambiguous areas caused by low texture and occlusion.

Global methods are based on the piecewise continuous constraint and coherence constraint. These algorithms realize the global optimal matching by minimizing the energy function through different ways, such as, dynamic programming (Birchfield and Tomasi, 1999), belief propagation (Klaus et al., 2006), and graph cut (Hong and Chen, 2004; Kolmogorov and Zabih, 2001), MicMac (Pierrot-Deseilligny and Paparoditis, 2006). Some semi-global methods also exist, such as semi-global matching (SGM) (Hirschmüller, 2005; Hirschmüller, 2008) and matching cost with a convolutional neural network (Žbontar and LeCun, 2015), which minimize the energy function through the cost aggregation of multi-directions. The common features of global and semi-global methods are their adoption of various constraints in the procedure of cost matrix calculation and their usually time- or memory-consuming minimization of the energy function (e.g., graph cut based method and SGM, respectively). Thus, these methods are not capable of directly operating on large images, unless supplemented with additional procedures, such as partitioning, which leads to other problems, and cannot be used in practical surveying and mapping production.

In particular, most of the existing approaches (1) depend on the geometric information of photography, especially, epipolar geometric information, (2) may be influenced when the camera has severe distortion, and (3) cannot automatically adapt to the variances in sensor types (from frame to line array) and imaging types (from optical to radar images).

In this paper, we introduce a method based on two steps of expansion and inspired by the seed-and-grow method (Adams and Bischof, 1994; Sharma et al., 2011; Wang et al., 2013; Yin et al., 2014). It determines the corresponding information on all the feature points of an image through two steps of expansion over a small number of seed and feature points. It also exhibits good performance over the area of disparity discontinuities. Given the idea of expansion, our method can easily deal with two frames without epipolar constraints. In other words, our method

---

does not need exterior and interior parameters of images, which in turn makes our method operate smoothly on frame aerial, line array, synthetic aperture radar, and natural scene images (Haala, 2013; Cavegn et al., 2014; Sharma et al., 2014). Our method also guarantees fast calculation speed.

2. FORMULATION

In this section, we describe the proposed integrated approach step by step and explain why we choose the Harris corner detector (Azad et al., 2009) as the feature extraction method, scale invariant feature transform (SIFT) matching (Lowe, 1999; Lowe, 2004) as the seed points matching method and normalized cross-

correlation (NCC) (Tsai and Lin, 2003) as the matching cost used in the expansion step of our approach.

Figure 1 illustrates the workflow of our two steps of expansion (TSE) method somewhat similar to patch-based multi-view stereo (PMVS) (Furukawa and Ponce, 2010). First, the seed points are extracted from separately extracted features of the base and the matching image using the feature matching method (e.g., SIFT matching) and are then used to estimate the location and size of the pre-matching patch (i.e., the first and the second step of expansion). Finally, the traditional correlation-based method is used to obtain the final refined matching.
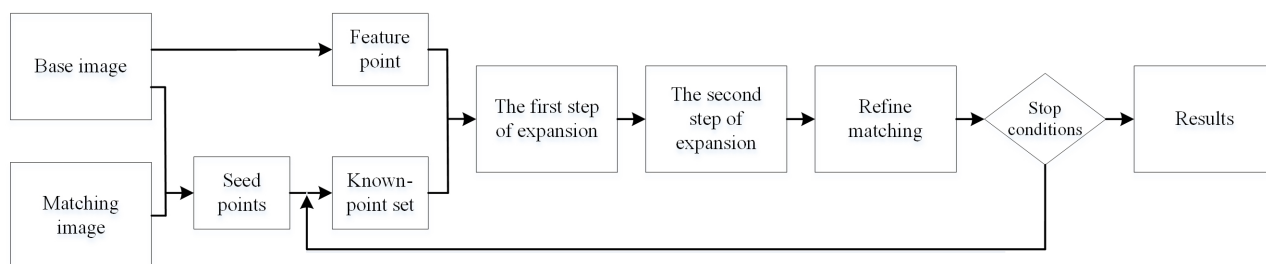

Figure 1. Pipeline of the proposed method

2.1 Extraction of Seed and Feature Points

Seed points are used to estimate the initial location and parallax search range of the possible matches. Although our method does not rely on the location accuracy of seed points (which is explained later), we still need to find a robust approach that can easily handle changes in illumination, rotation, zoom, and other such conditions. Furthermore, the efficiency of the approach is also required, which is important in considering the overall computing efficiency of our method. Only a small number of seed points are needed, so that the feature matching method is the first choice.

The feature points, whose parallax needs to be determined, are extracted only from the base image. The Harris corner detector is a suitable choice for calculating feature points, given its capacity to extract feature points from windows of various sizes (e.g., 3×3, 5×5, and 9×9), which is important because the final dense matching result is a subset of the feature points. In other words, the more the feature points, the denser the final matching result.

Many popular and excellent feature matching methods are available, including SIFT matching, gradient location-orientation histogram matching (Mikolajczyk and Schmid, 2005), and speeded up robust feature matching (Bay et al., 2008), to name a few. SIFT (1) is invariant to rotation, scale changes, and affine transformations, (2) has good performance in handling occlusion and complex backgrounds, and (3) has near real-time performance (Juan and Gwun, 2009). Thus, this method is used to find seed points to ensure robustness in dealing with various types of images.

In this part, SIFT matching is used to obtain the initial matching points, that is, the seed points (i.e., the triangle shown in Figure 2, marked as P_{seed}), which form the seed point setand the initial known-point set (i.e., the square shown in Figure 2, marked as P_{known}). The feature points are extracted from the base image using the Harris corner detector at certain intervals (cx), and they form the feature point set that needs to be matched (i.e., the cross shown in Figure 2, marked as P_f). The number of seed points

must exceed a given threshold (empirically set to 10 in our experiments); otherwise, our method may fail. The seed points should homogeneously distributed on the base image, otherwise the expansion speed will be affected.
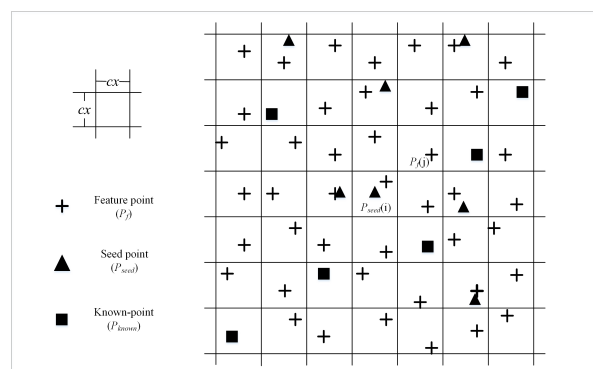

Figure 2. Example of the possible distribution of the feature, seed, and known points on an image. cx denotes the size of the Harris feature extracting window.

2.2 The First Step of Expansion

The first step of expansion is designed to find all the P_f around the P_{known} in the known-point set. The expansion can be defined as in the dilation of morphology:

$$D(X) = \{x | B[x] \cap x \neq \varnothing\}$$
$$= X \oplus B$$

(1)

where $x =$ the P_{known} in the first step of expansion or the P_f in the second step of expansion and belongs to X

$B[\ ] =$ denotes the operation of the growth of eight neighbours

$D(X) =$ refers to the P_f found in the first step of expansion or the P_{known} found in the second step of expansion

Taking $P_{seed}(i)$ (Figure 3) as an example, this procedure can be described as follows:

Assuming that this is the first time to take the first step of expansion on $P_{seed}(i)$, then, along the expansion route no. 1, we can find seven feature points that need to be matched, including $P_f(j)$ and $P_f(1), P_f(2), P_f(3), P_f(4), P_f(5), P_f(6)$ .The expansion route no. 2 must be searched when $P_{seed}(i)$ is processed for the second time, the expansion route no. 3 for the third time, and so on. This procedure of a known point is stopped when (1) the search route encounters the border of a base image or (2) the search route of two neighbor known points run into each other. Note that the known-point set then changes in every iteration.
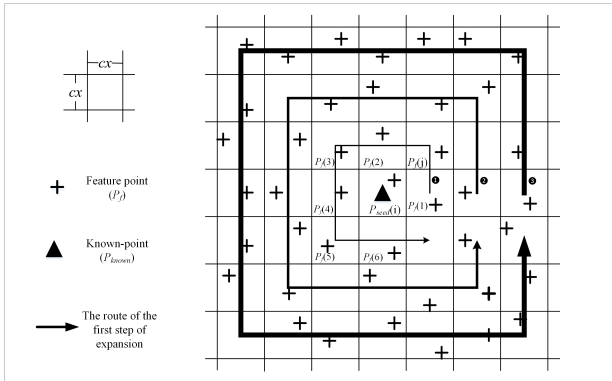

Figure 3. First step of expansion. The size of the route indicates a different iteration. For example, the expansion route no.1 is used at the first time of iteration, the expansion route no.2 at the second, and so on. The known point in the figure happens to be a seed point and is therefore represented by a triangle.

2.3 The Second Step of Expansion

In the second step of expansion, several known points near the feature point are found and used to calculate the initial location and size of the possible matching area of the feature point through "means of power of distance," which can be defined as follows:

$$x_{p'} = x_p + \sum_{(q_i, q'_i) \in Q} \left((x_{q'_i} - x_{q_i}) \times \left(1 - \frac{S_{p,q_i}}{\sum_{(q_j, q'_j) \in Q} S_{p,q_j}} \right) \right) \quad (2)$$

$$y_{p'} = y_p + \sum_{(q_i, q'_i) \in Q} \left((y_{q'_i} - y_{q_i}) \times \left(1 - \frac{S_{p,q_i}}{\sum_{(q_j, q'_j) \in Q} S_{p,q_j}} \right) \right)$$

$$S_{p,q} = \sqrt{(x_p - x_q)^2 + (y_p - y_q)^2}$$

$$i, j = 1, 2, 3, ..., N$$

where N = the number of known points that are found around Pf and that form the set Q

$p(x_p, y_p)$ = the coordinate of the feature point on the base image

$p'(x_{p'}, y_{p'})$ = the initial location of the possible match point of P_f on the matching image

$(q_i(x_{q_i}, y_{q_i}), q'_i(x_{q'_i}, y_{q'_i}))$ = an element of Q specifically, the coordinate of a matching point pair (on the base and matching images)

An example on $P_f(j)$ is given (as shown in Figure 4) to further explain the second step of expansion. We can find eleven known points along the expansion route shown in Figure 4,

including, $P_{seed}(1), P_{seed}(2), P_{seed}(3), P_{seed}(4), P_{seed}(5),$ $P_{seed}(6); \quad P_{known}(1), P_{known}(2), P_{known}(3), P_{known}(4),$ $P_{known}(5), P_{known}(6)$ and $P_{seed}(i)$. The initial location (i.e., $p'(x_{p'}, y_{p'})$) of the possible matching area of $P_f(j)$ is calculated using the "means of power of distance." The initial size (i.e., $Rang(j)$) of the possible matching area is determined by the bounding rectangle of the coordinate of the matching point pair in Q. This procedure accounts for the relationships between the feature and known points and therefore enhances the robustness of the matching. N should exceed a particular threshold, which is empirically set to 10 in our experiments.
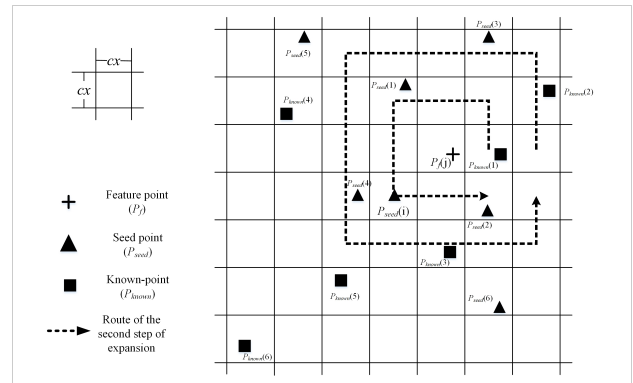

Figure 4. Second step of expansion

2.4 Refine Matching

After the second step of expansion, we obtain the initial match that is refined in this step. We use the correlation-based local method to refine the match. A very significant reason behind this choice is the computational efficiency of local methods. We choose NCC as our matching cost because it (1) compensates for the differences of pixels (in the window) in gain and bias, (2) considers intensity changes, and (3) is statistically optimal in compensating for Gaussian noise (Einecke and Eggert, 2010), although it is not robust in the area of low or repeated texture.

Pyramid matching is adopted in this procedure, which works as follows: First, the pyramid images of the base and matching images are built around the center of the initial match. Next, intensity correlation matching is then conducted from the top to the bottom of the pyramid image. Finally, the refined match on the origin layer of the base and matching images is found. In other words, we obtain the best match point ($Q_f(j)$) of $P_f(j)$ and the correlation coefficient ($C_f(j)$), which is used to evaluate the match.

After every iteration of P_{known} we conduct a statistical analysis of the correlation coefficient that is grouped by the known point. If the percentage of the matched feature points (i.e., those found around a known point) whose correlation coefficient is below a given threshold (*Thresh_C*, commonly set to 0.8) is greater than a given threshold (*Thresh_P*), then the known point is considered to be unreliable and should be removed from the known-point set. This process reduces the reliance on the accuracy of seed points but improves the robustness of our method.

If a known point is considered to be unreliable, the group of matches expanded from this known point is also unreliable and should be abandoned. Otherwise, the matching pair whose correlation coefficient exceeds a given threshold (*Thresh_C*) is categorized in the known-point set. A new iteration of the two steps of expansion is then executed until the number of new

known points found in the last iteration is below a certain threshold (*Thresh_N*).

When the main matching process ends, the density analysis method (Zhu et al., 2015) is used to eliminate large outliers, after which the final matching results are obtained.

The choice of the method mainly depends on the application. In other words, other methods can be used to find seed points (e.g., correlation based method on thumbnails of base and matching image), as well as other matching costs in the procedure of refining the matching.

3. EVALUATION

To evaluate the proposed approach, five types of images are used in our experiments. The experimental results of our proposed method are given, after a brief introduction of the images used in our experiments. The size of the Harris feature extractor window is set to 3 × 3 to make the matching results as dense as possible. All the experiments are conducted on a portable computer with Intel Core i3-3110M CPU and 2.4GHz. And we empirically set the *Thresh_C* to 0.8, *Thresh_P* to 60%, and *Thresh_N* to 100.

3.1 Description of the Test Images

We use two pairs of aerial frame images from Vaihigen and San Francisco; one pair of images captured by ADS40; one pair of satellite images by GeoEye-1; one pair of single-polarization SAR images; and test images (specifically, teddy, cones, and venus) from the Middlebury benchmark (Scharstein and Szeliski, 2002). The details of the frame images are shown in Table 1.

Descriptions/ Stereo Pair	Stereo Pair 1	Stereo Pair 2
Location	San Francisco (USA)	Vaihigen (Germany)
Camera	UltraCam D	DMC
Image size(pixel)	11,500 × 7500	7680 × 13,824
Side and Forward Overlap	30/60%	60/60%
GSD	15 cm	8 cm
Description of the image area	Densely-distributed buildings	Small-sized and detached buildings surrounded with many trees

Table 1. Overview of the two pairs of frame images

The first pair of line array images is cut from the nadir and backward strip of the airborne ADS40, which is a common line array camera. The pair of images is captured from suburban areas. The second pair of line array images is cut from a panchromatic image of GeoEye-1. Many low-rise buildings are present in the second pair of images.

The SAR image pair is a same-side stereo pair, with a GSD of 0.5 m, and a central angle of incidence of 45°. The image area of this pair is a hilly area, and a small town exists in the area.

3.2 Results

3.2.1 Frame Image: The two pairs of images both have accurate exterior and interior parameters, and the stereo pair from Vaihigen (as short as S2) has LiDAR points. The dense 3D point cloud of the stereo pair from San Francisco (as short as S1) is generated using SGM. The result of SGM is down-sampled to a

3 ×3 pixel. Our results are also transformed to the object space coordinate system through forward intersection.

To compare the result of S1 with SGM and the result of S2 with LiDAR, checkpoints are selected randomly, and differences in elevation with the corresponding points in the SGM results (S1) or LiDAR (S2) of the check points are used as the assessment index.

Figure 5 shows the result of S1, and Figure 6 shows the result of S2. Given that NCC is used to refine the match, our method fails or results in a blur in low-texture (as shown in Figure 5(e)), and repeated texture areas (as shown in Figure 6(f)). According to the profile, our result has a similar precision with SGM. The density of the point cloud obtained with our method is dramatically higher than that of LiDAR and at the same time has a near precision. Similar to the other matching methods, our method can obtain matching results on walls, which is difficult for LiDAR.
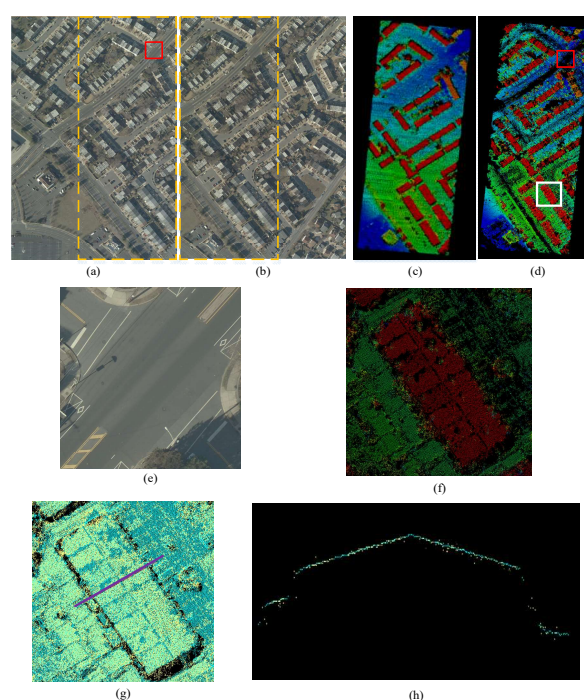

Figure 5. Matching result of S1 and comparison with SGM, where (a) and (b) are the base and matching images of the stereo pair, respectively, and the areas within the dashed boxes represent the overlapping of the image pair; (c) is the result of SGM (coloring by elevation of the point, the higher the redder, otherwise, the lower the more blue); (d) is the result of our method (the coloring is the same as (c)); (e) is the enlarged view of the area in the red rectangle of (a) and is also the base image of the unmatched area in the red rectangle of (d); (f) is the enlarged view of the area in the white rectangle of (d); (g) is the enlarged view of the same area of (f), but the SGM result (the blue point) and our result (yellow point) are put together; (h) is a profile of the purple line in (g).

The statistical results of the elevation differences of the check points are given in Table 2. The maximum absolute elevation difference of S1 and S2 are 9.5 and 10.3m, respectively. The percentages of the points whose absolute value of elevation difference is less than 0.5m in our results are more than 93% (95% in S1 and 93.5% in S2). Mismatching appears on (1) the roof of a building, which generally has a repeated texture, (2) the surface of the road, on which little information exists, and (3) the area

with amorphous trees. Table 3 shows the matching time of S1 using SGM and our method. Our method has a similar accuracy with SGM, costs less time than SGM, and is specifically 4.2 times faster than SGM. However, our result is not a pixel by pixel result.
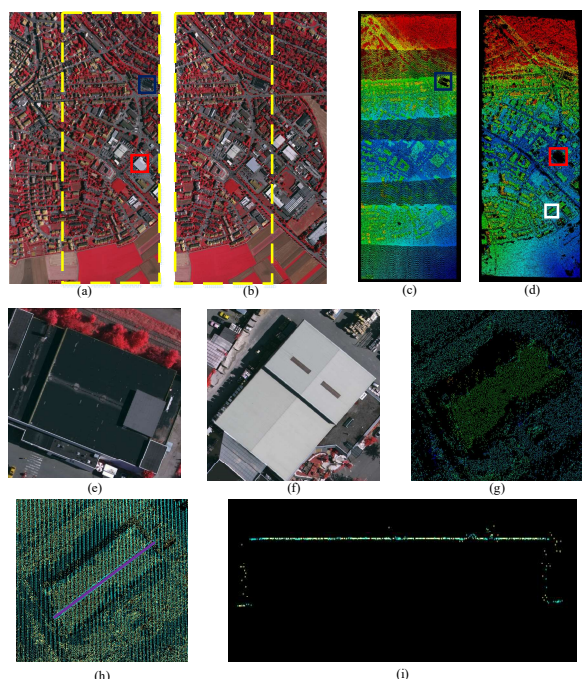

Figure 6. Matching result of S2 and comparison with LiDAR, where (a) and (b) are the base and matching images of the stereo pair, respectively, and the areas within the dashed boxes represent the overlapping of the image pair; (c) is the point cloud of LiDAR (the coloring is the same as that in Figure 5(c)); (d) is the result of our method (the coloring is the same as that in Figure 5(c)); (e) is the enlarged view of the area in the blue rectangle of (a), which has no LiDAR point (as shown in the blue rectangle in (c)); (f) is the enlarged view of the area in the red rectangle of (a) and is also the base image of the unmatched area in the red rectangle of (d); (g) is the enlarged view of the area in the white rectangle of (d); (h) is the enlarged view of the same area of (g), but the LiDAR point (the blue point) and our result (yellow point) are put together; (i) is a profile of the purple line in (h). (Note: The LiDAR is captured in strip, so that a strip change is evident in (c)).

Stereo pair		S1	S2
Number of matching point		2,329,580	3,681,388
Number of check point		1,165,524	2,046,240
Absolute value of the elevation difference (m)	[0, 0.5)	1,107,247	1,913,629
	[0.5, 2)	36,154	71,853
	[2, ∞)	22,123	60,758

Table 2. Statistic results of elevation difference of the frame images.

Time/Method	Our Method	SGM
Processing Time (s)	242.699	1,267.985

Table 3. The processing time of S1 (Note: the time of SGM including the time of generating epipolar image)

3.2.2 Line Array Image: Every single line of the line array image has different exterior parameters, which means it is troublesome for algorithms using epipolar constraints to match. Two pairs of line array images are used to test the ability of our method to match this type of image. The base image of the first

pair is cut from the nadir strip image, and the matching image is cut from the backward strip image. The second pair is cut from a GeoStereo product of GeoEye-1. To visualize our results, we use the parallax as the Z value and the coordinate on base image as the X and Y values to form a point cloud. The color change indicates a parallax change. As shown in Figure 7(c), the points of the buildings are red, which means these points have a large parallax. By contrast, the points in flat areas are almost blue, which means these points have a small parallax. Information on the local details, specifically, the buildings, is lost in our results because low-rise buildings have a very small parallax change on the satellite image. However, our result evidently represents the terrain correctly. In other words, our method is capable of matching line array images.
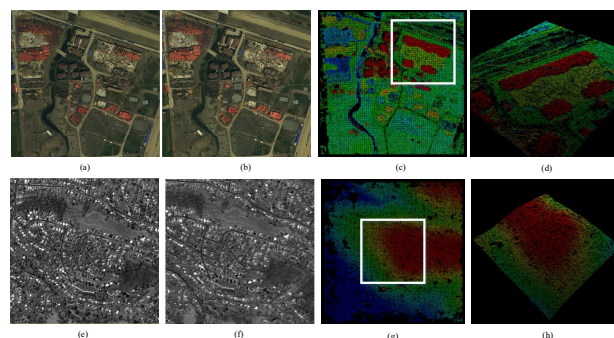

Figure 7. Matching result of the line array image, where (a) and (b) are the base and matching images of the stereo pair from ADS40; (c) is the result of (a) and (b) (coloring by parallax, red represents a large parallax, and blue represents a small parallax); (d) is the enlarged view of the area in the white rectangle of (c) shown in 3D; (e) and (f) are the base and matching images of the stereo pair from GeoEye-1; (g) is the result of (e) and (f) (the coloring is the same as (c)); (h) is the enlarged view of the area in the white rectangle of (g) shown in 3D.

3.2.3 SAR Image: The imaging apparatuses of SAR differ from those of an optical image. Repeated texture appears frequently on SAR images (as shown in Figure 8(d)). The scale of a SAR image changes along with the distance between the target and the nadir point of the sensor. In other words, the size and the shape of an object differ on stereo (which can be shown by the different sizes of the dashed box in Figures 8(a) and (b)), and the corresponding information is therefore difficult to be found. Figure 8(c) shows the parallax changes from left to right, which are correct according to the imaging apparatuses of SAR.
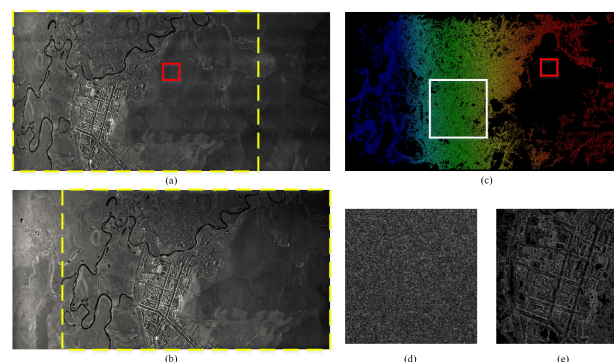

Figure 8. Matching result of S2 and comparison with LiDAR, where (a) and (b) are the base and matching images of the stereo pair, respectively, and the areas within the dashed boxes represent the overlapping of the image pair; (c) the result of our

method (coloring by parallax, red represents a large parallax, and blue represents a small parallax); (d) is the enlarged view of the area in the red rectangle of (a) and is also the base image of the unmatched area in the red rectangle of (c); (e) is the enlarged view of the area in the white rectangle of (c), but the coloring is by the grey value of the base image.

3.2.4   Natural Scene Image: We also test our method on the Middlebury benchmark. To ensure the matching robustness, we use the Harris feature point as the feature point, which unfortunately is incapable of pixel-by-pixel extraction. In other words, our method cannot generate a dense depth map. The statistic for the depth differences of our matching point with the ground truth is used to evaluate our results, although it is not calculated pixel by pixel. The details are given in Table 4.

We set a high threshold of correlation coefficient to ensure the accuracy of our method. The percentages of the wrong match in our results are less than 20%. The mismatch generally appears in the area of the repeated texture. Considering the processing time of our method, our results can be used to provide initial depth information on other pixel-by-pixel methods.

Data		Teddy	Cones	Venus
Image size (pixel)		450 × 375	450 × 375	434 × 383
Processing time (s)		1.687	1.862	1.463
Number of corresponding points		10461	10844	9327
Max. of difference of depth (MDD) (pixel)		8	14	5
Depth difference (pixel)	0	8627	9057	8475
	1~2	1784	1727	605
	3~5	47	58	247
	6~MDD	3	2	/

Table 4. Statistical results of processing time and depth difference of the natural scene images

4.   CONCLUSION

A dense matching method with two steps of expansion, which can be categorized as a seed-and-grow method, is proposed in this paper. In this method, matching results of the feature-based method are used as seed points and as the initial known-point set. The first step of expansion is conducted on the known-point set iteratively to find the feature points. The second step of expansion is conducted on the feature points found in the first step of expansion, and the initial match area is found after two steps of expansion. Finally, the result is refined using the correlation-based method. This method needs no prior information, especially the geometric information of an image, which makes this method applicable to various types of images. The result of this method has high accuracy and can be used as the final dense matching result to some extent or the initial value for other dense matching methods that aim to generate pixel-by-pixel point cloud. Moreover, this method saves time.

However, the proposed method is not capable of dealing with large areas of low or repeated texture because of the matching cost we used in the refined matching step, NCC, is not robust enough in handling these ambiguous areas. This method also fails when no seed point is available. The following issues need to be explored in future works:
1) A matching cost that is robust in large ambiguous areas of low or repeated texture

2) More robust and efficient model used in estimating the initial location and size of possible matching areas, which can address large distortions and uneven distribution of seed points
3) More efficient filtering algorithm for generating more accurate results

REFERENCE

Adams, R. and Bischof, L., 1994. Seeded region growing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(6), pp. 641-647.

Azad, P., Asfour, T. and Dillmann, R.U.D., 2009. Combining Harris interest points and the SIFT descriptor for fast scale-invariant object recognition. In: *IEEE Computer Society Conference on Intelligent Robots and Systems (IROS)*, St. Louis, MO, USA. IEEE, pp. 4275-4280.

Bay, H., Ess, A., Tuytelaars, T. and Van Gool, L., 2008. Speeded-up robust features (SURF). *Computer Vision and Image Understanding*, 110(3), pp. 346-359.

Birchfield, S. and Tomasi, C., 1999. Depth discontinuities by pixel-to-pixel stereo. *International Journal of Computer Vision*, 35(3), pp. 269-293.

Cavegn, S., Haala, N., Nebiker, S., Rothermel, M., and Tutzauer, P., 2014. Benchmarking high density image matching for oblique airborne imagery. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40(3), pp. 45.

Einecke, N. and Eggert, J., 2010. A two-stage correlation method for stereoscopic depth estimation. In: *IEEE Computer Society Conference on Digital Image Computing: Techniques and Applications (DICTA)*, Sydney, Australia. IEEE, pp. 227-234.

Furukawa, Y. and Ponce, J., 2010. Accurate, dense and robust multiview stereopsis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(8), pp. 1362–1376.

Haala, N., 2013. The landscape of dense image matching algorithms. *Photogrammetric Week '13 (Ed. D.Fritsch)*, Wichmann, Berlin/Offenbach, Germany. pp. 271–284.

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 (CVPR)*, San Diego, CA, USA. IEEE, pp. 807-814.

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

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

Juan, L. and Gwun, O., 2009. A comparison of sift, pca-sift and surf. *International Journal of Image Processing (IJIP)*, 3(4), pp. 143-152.

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

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

Lowe, D.G., 1999. Object recognition from local scale-invariant features. In: *IEEE International Conference on Computer Vision (ICCV)*, Kerkyra, Greece. IEEE, pp. 1150-1157.

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

Mikolajczyk, K. and Schmid, C., 2005. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10), pp. 1615-1630.

Pierrot-Deseilligny, M., and Paparoditis, N., 2006. A multiresolution and optimization-based image matching approach: An application to surface reconstruction from SPOT5-HRS stereo imagery. *Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36(1/W41).

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

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

Sharma, A., Horaud, R., Cech, J. and Boyer, E., 2011. Topologically-robust 3D shape matching based on diffusion geometry and seed growing. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Colorado Springs, CO, USA. IEEE, pp. 2481-2488.

Tsai, D. and Lin, C., 2003. Fast normalized cross correlation for defect detection. *Pattern Recognition Letters*, 24(15), pp. 2625-2631.

Wang, G., Yin, X., Pei, X., and Shi, C., 2013. Depth estimation for speckle projection system using progressive reliable points growing matching. *Applied optics*, 52(3), pp. 516-524.

Yin, X., Wang, G., Shi, C., and Liao, Q., 2014. Efficient active depth sensing by laser speckle projection system. *Optical Engineering*, 53(1), pp. 013105-013105.

Žbontar, J. and LeCun, Y., 2015. Computing the stereo matching cost with a convolutional neural network. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA. IEEE, pp. 1592-1599.

Zhu, J., Hu X., Zhang, Z., , Xiong, X., 2015. Hierarchical outlier detecion for point cloud data using a Density Analysis Method. *Acta Geodaetica et Cartographica Sinica*, 44(3), pp. 282-291.