The Performance of the Optical Flow Field based Dense Image Matching for UAV Imagery

Wei Yuan[1], Weihang Ran[2], Bruno Adriano[1], Ryosuke Shibasaki[2], Shunichi Koshimura[1]

[1]International Research Institute of Disaster Science, Tohoku University, 980-8572 Sendai, Japan
{wei.yuan, bruno.adriano, koshimura}@tohoku.ac.jp
[2]Center for Spatial Information Science, the University of Tokyo, 277-8568 Kashiwa, Japan
{ranweihang, shiba}@csis.u-tokyo.ac.jp

Keywords: Dense Image Matching, UAV Imagery, 3D Reconstruction, Quality Assessment, Sustainable Development Goals.

Abstract

With the rapid development of sensing platforms, unmanned aerial vehicle (UAV)-based mapping has become increasingly popular because of its economic efficiency and flexibility, especially for providing 3D information to support urban growth monitoring and change detection to meet sustainable development goals (SDGs). This paper presents an improved optical flow field-based dense matching algorithm (OFFDM) for low-altitude UAV images based on the Ph.D. thesis of Yuan (Yuan, 2018). First, high-precision seed points were used to compute the optical flow field within stereo pairs, effectively minimizing redundant calculations during the fine-matching phase. Second, a fine-matching approach, integrating multiple constraints, was applied to refine the coarse matching results based on the optical flow field. Extensive dense matching experiments on UAV low-altitude aerial imagery assessed the performance of OFFDIM across four dimensions: 3D point cloud visualization, matching success rate, precision, and reliability. Extensive experiments on low-altitude UAV imagery, characterized by a resolution of 7cm per pixel over a 10,608x8,608 pixel dimension and a 60% forward overlap, evaluate the OFFDM's efficacy. The quantitative evaluation revealed that the proposed method achieved an accuracy of ±0.7 pixels in image coordinates and ±20 cm on the ground, with a matching success rate exceeding 97%. The processing time was approximately 272 seconds for handling one single stereo pair. When compared to the widely adopted PMVS algorithm, known for its effectiveness in dense matching for UAV images, the proposed method demonstrated higher completeness and improved matching efficiency by more than five times. These results demonstrated that the proposed approach is more suitable for dense matching on UAV imagery-based high-precision 3D spatial data extraction, supporting global mapping tasks more effectively.

1. Introduction

In recent years, unmanned aerial vehicles (UAV) images have been widely used for information extraction in three-dimensional geographic spatial information. Due to advantages such as low flying altitude, convenient operation, relatively lenient weather requirements, and cost efficiency, the application scope of UAVs has been expanding rapidly (Turner, 2012), which will meet sustainable development goals of continuous mapping of the changing world. However, due to limitations such as small mass, poor flight stability, and significant susceptibility to airflow, UAV images often exhibit large viewing angle shifts and illumination variations, making it difficult for dense image matching (Rock, 2011; Xiao, 2016).

Traditional dense image-matching methods are usually based on determining the best matching cost for correspondence searching, the matching costs are calculated by pixel intensity through the Markov random field. However, these methods usually get into local optima. Some other improved methods utilize multi-view images for adding geometric correlations between co-viewed images, which may make the matching results more accurate and robust to occlusions, but are computationally expensive. Nowadays, with the rapid development of deep learning techniques, deep learning-based stereo-matching, and dense image matching methods are dominant in almost all the open benchmarks. However, most of them are supervised learning-based methods, which require highly accurate ground truth data for training the neural network parameters. For real applications, acquiring large amounts of ground truth data is expensive and time-consuming; on the other hand, domain shifts make the pretrained neural network fail to deal with unseen scenes, which makes deep learning-based methods unfeasible.

To solve the above difficulties, this paper aims to enhance the previously proposed optical flow field-based dense image matching method (Yuan, 2019). It seeks to evaluate its performance both subjectively and quantitatively across four key aspects: visual matching effect, matching success rate, matching reliability, and matching accuracy. The goal is to enhance its suitability for the dense matching of low-altitude aerial images captured by UAVs. The comparative experiment against the widely recognized Patch-based multi-view stereo matching (PMVS) method further explores the potential application of the optical flow field-based dense image matching method in generating high-precision dense 3D point clouds from UAV imagery to support the global mapping downstream tasks.

2. Related Works

Dense image matching has received substantial interest from both photogrammetry and computer vision fields and has seen significant advancements over the past few decades (Rothermel, 2011; Remondino, 2014). Whereas in photogrammetry, the stereo image matching strategy is mainly used. The task can be summarized as identifying the position of corresponding image points in a stereo pair by accounting for both geometric and radiometric constraints between the images (Torresani, 2013). According to the different methods of matching cost optimization, dense image matching methods can be further classified into local optimization methods (Ke, 2004;) and global optimization methods (Issac, 2014; Tran, 2006;). The local optimization method employs the Winner-Takes-All (WTA) strategy, which selects the final correspondence with the lowest matching cost. By calculating the matching cost between the target point and its neighbourhoods the correspondence is selected (Tola, 2008). The advantage of this method is that the computational complexity is low and the amount of redundant computation is small, but it is easy to fall into the local optimum, which leads to the matching result not in line with the actual

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-4-2024-433-2024 | © Author(s) 2024. CC BY 4.0 License.

terrain; the global optimization method is to optimize the cost of matching based on pixels or objects by constructing a global optimization energy function, so as to make the final matching result reach the global optimum (Hirschmüller, 2009). This method has high matching accuracy because it takes into account the information of every pixel in the image, but it has more redundant computations and is not efficient. Despite the above advantages of the stereo image matching method, its matching results are still less robust to occlusion and noise because only the information on the two images is taken into account when matching (Remondino, 2008).

In computer vision, multi-view matching methods have become a prominent research topic (Seitz, 2006). By incorporating geometric relationships and redundant information across multiple images, these methods produce matching results that are significantly more resilient to occlusion and noise compared to traditional stereo image matching algorithms. One widely adopted approach is the PMVS method, introduced by Furukawa and Ponce (2010). PMVS requires no prior knowledge or initialization settings and is well-suited for 3D reconstruction of large scenes, making it a popular choice in UAV low-altitude photogrammetry. Ai et al. (2015) enhanced PMVS by introducing high-precision sparse matches as seed points for dense matching in UAV images, significantly improving the method's efficiency. Additionally, Shao et al. (2016) leveraged PMVS’s results as initial values, constructing an expanded patch set and refining the positions of matching points within the patch using least squares refinement and MPGC (Baltsavias, 1996), leading to more robust matching results under occlusion and noise, and generating a denser point cloud.

In recent years, deep neural networks have demonstrated exceptional performance in object detection and classification tasks. Due to their robust feature extraction capabilities, convolutional neural networks (CNNs) have also been extensively applied to optimize dense image matching costs. In 2015, Zbontar and LeCun pioneered the use of CNNs for patch-based dense image matching, framing the search for dense correspondences as a binary classification problem to identify pixel-wise matches using deep networks. Subsequent advancements have centered on improving network architecture, with significant contributions such as MatchNet (Han, 2015). Researchers have further sought to optimize stereo depth estimation within constrained size and matching ranges by designing specialized architectures and loss functions. GCNet (Alex, 2017) introduced a technique for generating 3D cost by comparing features of reference image pixels with potential matching pixels in the target image. Similarly, PSMNet (Chang, 2018) utilizes pyramid space pooling and hourglass networks to effectively capture image context.

3. Method

3.1 Improved OFFDIM

The proposed approach includes three steps: acquisition of sparse optical flow field, estimation of dense optical flow field, and point cloud refinement, which is identical to the PMVS method in the basic idea. The traditional PMVS method first acquires seed points through feature matching, constructs matching patches based on seed points, expands seed patches to obtain dense patches, and finally filters and optimizes dense patches to obtain dense matching points. A detailed description of the OFFDIM proposed by Yuan (2019) will be given in this paper.

3.1.1 Acquisition of Sparse Optical Flow Field

As traditional optical flow fields cannot track the large-amplitude motion of pixels between image sequences, sparse optical flow fields between images were obtained in this paper through pyramid L-K optical flow field (Bouguet, 2001) based on feature points. Concrete steps were: (1) a pyramid image between two images to be matched was established, (2) optical flow fields of feature points on the top pyramid image were tracked, and (3) sparse optical field fields on the raw image were acquired through iterations. The advantages of pyramid L-K optical flow field were: high robustness, pyramid L-K optical flow field was of favorable robustness to image noise, brightness nonuniformity and some covers and obtained sparse optical flow fields had high precision; high computation speed, search strategy based on image pyramid could significantly reduce search range of optical flow and improve overall operating rate.

3.1.2 Estimation of Dense Optical Flow Fields

Estimation of dense optical flow fields aims to obtain the approximate motion vector of the stereoscopic image for each pixel in the overlapping region and take it as the initial value for point cloud optimization. This can significantly reduce the redundant operation of searching for corresponding image points. After sparse optical flow fields are obtained, the construction of the optical flow of the stereoscopic image for each pixel in the overlapping region can be regarded as a process of interpolating unknown points from the discrete point set. The detailed dense optical flow fields estimation is illustrated in our previous paper please check (Yuan, 2019) for details

3.1.3 Multi-constraints Correspondence Refinement

In the 1990s, Baltsavias (1991) put forward a multi-photo geometrically constraint matching method (MPGC) combining epipolar constraint and least squares image matching, and this method was widely applied in result optimization of all kinds of 3D reconstructions (Remondino, 2008; Rothermel, 2012; Torresani, 2013). As the convergence rate of least squares image matching is related to the accuracy of the given initial value and computation speed is relatively low, the refining process of a corresponding point pair (p, p′) obtained through a dense optical flow field is essentially judgment about whether it is located on corresponding epipolar line of the left and right images through epipolar constraint.

If (p, p′) is on the corresponding epipolar line, 15×15 pixel windows are selected on left and right images with p and p′ being the centers, respectively, and their normalized cross-correlation coefficient (NCC) is calculated. When it is greater than 0.85, the corresponding point pair (p, p′) does not need refinement. Otherwise, it is necessary to search out point q with the maximum NCC value with point p in the $\left(-\left|\vec{f}_{\max}-\vec{f}_{\min}\right|,\left|\vec{f}_{\max}-\vec{f}_{\min}\right|\right)$ search region along the corresponding epipolar line. When NCC_{pq} >0.6, (p, p′) is considered as the new corresponding point pair. Suppose (p, p’) is not on the corresponding epipolar line. In that case, it is necessary to find the corresponding epipolar line l corresponding to point p on the right image, and p′ should be projected onto *l* to obtain point s as shown in Fig. 1.

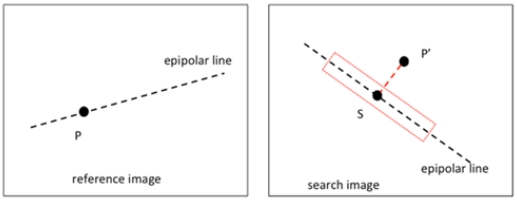


Figure 1. Definition diagrams of search region of corresponding points (red rectangular) (Yuan, 2018)

In consideration that orientation parameters between images may have errors, along the direction of epipolar line, the point *s′* with the maximum NCC value with point *p* is searched out in search region $(-\left|\vec{f}_{\max}-\vec{f}_{\min}\right|,\left|\vec{f}_{\max}-\vec{f}_{\min}\right|)$ on epipolar line *l* and in stripes of its upper and lower pixels with *s* being the center. When $NCC_{ps'}$, the corresponding point of point *p* on the right image is *s′*. According to the above operation, all pixels in the image overlapping region are transverse. After one pixel is checked, then the whole refinement process of the dense matching point is finished.

3.2 Quality Assessment of Dense Matching Point Clouds

In this section, we provide a comprehensive assessment of the quality of densely matched point clouds in two different aspects. The first step is the subjective visual inspection, we will visualize the 3D points clouds for detailed analysis. The second step is the quantitative analysis. The detailed analysis is shown as bellow

3.2.1 Matching Success Rate

In our experiments, each stereo pair is treated as a statistical unit. According to equation (1), the ratio of the total number of pixels in the overlapping region of the stereo images to the number of dense matching points obtained is calculated. This ratio reflects the effectiveness of the dense matching process in capturing corresponding points within the overlap area. A higher success rate indicates a more effective dense matching outcome (Wang, 2013).

$$MRS=\frac{\text{Number of the correspondence}}{\text{Number of overlap area pixels}}\times 100\% \quad (1)$$

3.2.2 Reliability

Based on the relative orientation elements, this paper calculates the stereo image pair's dense matching point cloud to calculate their vertical disparities individually. In the absence of mismatched points, the residuals of the vertical disparities for all points should conform to a normal distribution. According to Baarda's reliability theory, for a given point, when the significance level is set to 0.1%, the residual of the vertical disparities should not exceed 3.29σ, where σ represents the unit weight error in relative orientation (Li & Yuan, 2012). The dense matching point cloud is traversed using this criterion to count the out-of-limit points. Then, according to equation (2), the ratio of matching points to out-of-limit points within the stereo image pair is calculated to measure the reliability of dense matching. A smaller value indicates a higher reliability of dense matching.

$$r=\frac{\text{Number of out-of-limit points}}{\text{Number of correspondence}}\times 100\% \quad (2)$$

3.2.3 Accuracy in Image

Relative orientation involves analytically computing the intersection of corresponding rays in stereo image pairs, aiming to minimize vertical disparities in stereo models and assign reasonable image point coordinate observation errors. Since this process relies solely on image point coordinate measurements without incorporating non-photogrammetric data, the relative orientation unit weight error serves as an indicator of image point matching accuracy. However, given the vast number of dense matching points in each stereo pair, this paper employs continuous relative orientation through a weighted iterative method on seed points, removing outliers in the process. After calculating the relative orientation elements, the residual vertical disparities for each point in the dense matching point cloud are determined, and points with residuals exceeding 3.29σ are excluded. For the remaining points, the root mean square error (RMSE) of the residual vertical disparities is computed using equation (3), with a lower RMSE indicating greater dense matching accuracy.

$$\sigma=\sqrt{\frac{1}{n}\sum_{i=1}^{n}\Delta q_i^2} \quad (3)$$

In the equation, *n* is the number of dense matching points and Δq is a residual error of the vertical parallax of the *i*th orientation point.

3.2.4 Accuracy in Ground

After removing mismatched points, the dense matching point cloud is processed using the principle of stereo image intersection to calculate the three-dimensional ground coordinates of each point. This procedure results in a discrete 3D point cloud, commonly referred to as a Digital Terrain Model (DTM). Checkpoints derived from pass points using GPS-supported bundle block adjustment are then used to identify neighboring points in the DTM based on their planar coordinates. Since the optical flow-based dense matching achieves pixel-level correspondence, the nearest-neighbour interpolation (NNI) method is applied in this paper. This method determines the value of each point from the generated DTM closest to the checkpoint, allowing for the calculation of the elevation error. The accuracy of the DTM, denoted as *m*. *m* is calculated using equation (4), providing a measure of the in-ground accuracy of the matching point cloud. A smaller value indicates greater accuracy of the dense matching.

$$m=\sqrt{\frac{1}{n}\sum_{i=1}^{n}\Delta h_i^2} \quad (4)$$

In the equation, *n* is number of checkpoints and Δh is elevation error of the *i*-th checkpoint.

4. Experiments and Analysis

4.1 Datasets

The experiments were conducted using 44 unmanned aerial vehicle (UAV) images. The images were captured near the

Beijing area, primarily covering farmland, factories, roads, and small building areas. Table 1 presents the detailed parameters of the dataset. The highly accurate pass points were employed as checkpoints in the surveying area for the evaluation of the actual positioning accuracy.

Item	Beijing
Aerial craft	Unmanned Aerial Vehicle (UAV)
Camera	PhaseOne IXU-1000
Principal distance (mm)	51.21293
Format (pixels)	11608 × 8708
Pixel size (μm)	4.6
Ground sample distance (GSD) (cm)	7
Relative flying height (m)	779
Longitudinal overlap (%)	60
Lateral overlap (%)	30
Number of mapping strips	4
Number of control strips	4
Number of images	88
Number of ground control points	21
Number of pass points	55701
Block area (km2)	2.8 × 2.8
Maximum topographic relief (m)	54
Average terrestrial height (m)	508

Table 1. The parameters of the Beijing dataset

Figure 2 displays the distribution of 18 prominent ground control points (e.g., traffic lines, wall corners, road intersections, building edges) surveyed in the experimental area. These points function as ground control points, with their ground coordinates precisely determined through GPS static measurement methods. The coordinates achieved a field accuracy of within ±5 cm in all three directions, rendering them suitable for use as orientation points and checkpoints in photogrammetric measurements.

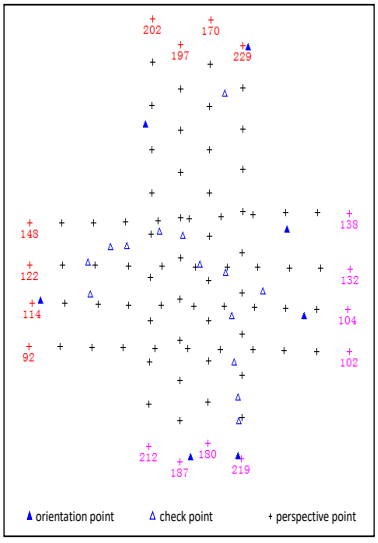
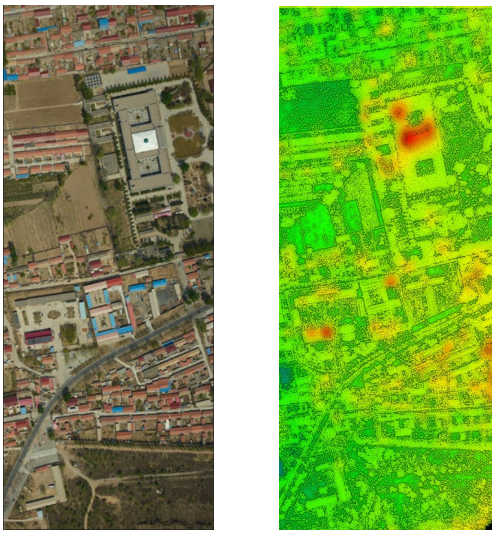


Figure 2. The strip plan of UAV datasets (Beijing)

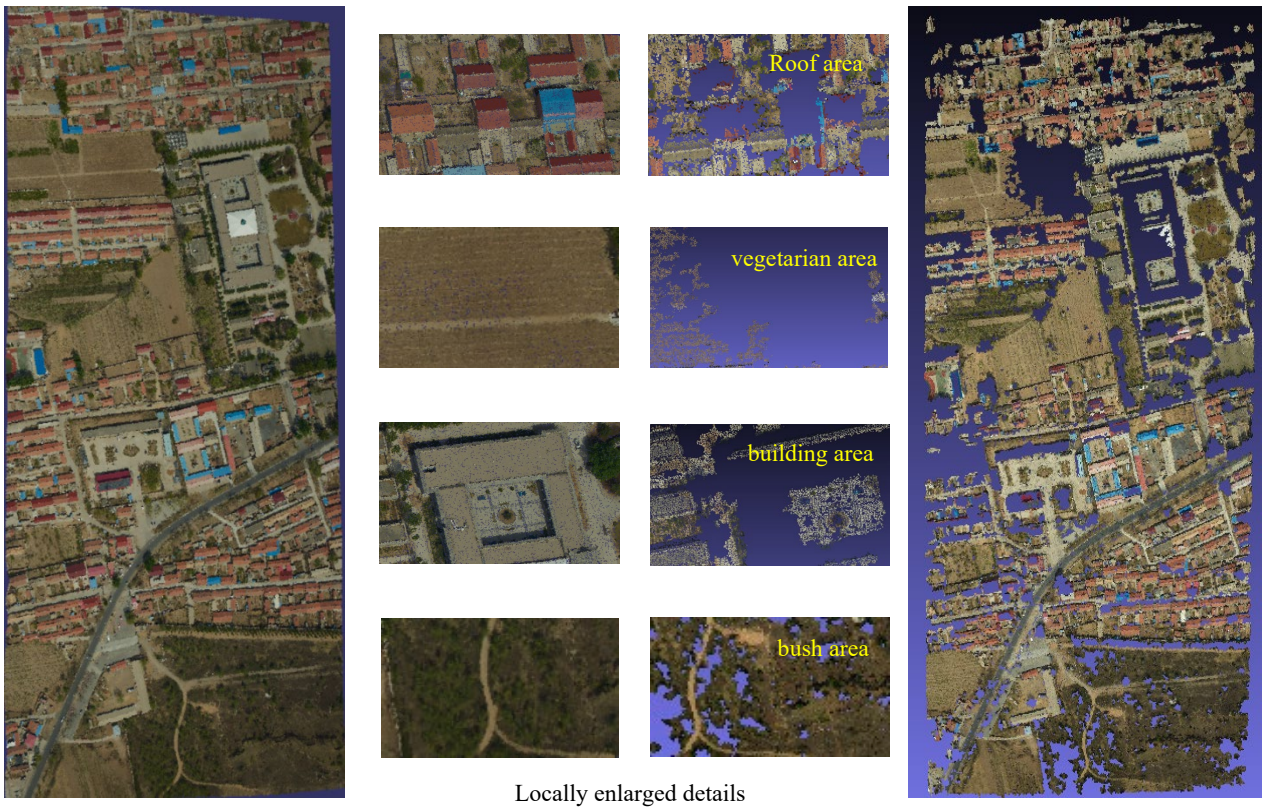
4.2 Dense Matching Effect

Given the extensive coverage and high feature repetition in the experimental images, we use stereo image pair 123-124, which encompasses a range of textures, to compare and subjectively evaluate the matching results of OFFDIM and PMVS. Figure 3(a) illustrates the overlapping area of the stereo images, featuring complex textures, including buildings, texture-deficient farmland, bare ground, and repetitive elements such as bushes, individual trees, and roads. Figure 3(b) presents the discrete Digital Terrain Model (DTM) point cloud automatically generated from OFFDIM’s dense matching results. The DTM accurately captures detailed features of various objects, with clear and complete edges for roads, buildings, and individual trees, consistent with the dense matching point cloud generated by OFFDIM in Figure 4(a). In contrast, Figure 4(b) shows the dense point cloud produced by PMVS for the same area. A comparison of Figures 4(a) and 4(b) reveals that OFFDIM generates a more complete point cloud, with fewer voids, especially in texture-poor farmland and areas with intricate textures (highlighted in the red box in the figure).



(a) Raw Image (b)Generated DTM

Figure 3. Dense matching effects (Yuan, 2018)



(a) OFFDIM generated point clouds (b) PMVS generated point clouds

Figure 4. Visual comparison of OFFDIM and PMVS (Yuan, 2018)

Upon closer inspection of these local areas, it is evident that OFFDIM can more comprehensively match the rooftops of buildings and obtain a denser matching point cloud for texture-poor farmland and grassland regions. Although the results of both methods are similar for road and bare ground areas, PMVS almost fails in bush and individual tree areas, while OFFDIM successfully matches complete point clouds. Thus, it is evident that the dense image-matching approach based on OFF fitting can yield denser and more complete matching point clouds for images with complex and texture-poor areas.

4.3 Dense Matching Quality Analysis

Firstly, the matching success rate of the three-dimensional point clouds within the overlap regions of 40 stereoscopic image pairs across 4 flight strips was calculated using Equation (1). The variation curve, in Fig 4, reveals that the matching-success rate exceeds 98.5%, with only three stereoscopic image pairs having a success rate of around 97.0%. This indicates that OFFDIM achieves high completeness in matching 3D point clouds within the overlap areas of stereoscopic image pairs. Additionally, the reliability curve of matching, as depicted in Figure 5, shows that the mismatching rate of the OFFDIM-generated dense point cloud is below 20 percent. Notably, the 140-141 pair, containing only 99 concentrated seed-points in local area, exhibited a high mismatch rate of up to 65%. Images in flight strip 148-138, which cover a large number of buildings with complex textures, also had a higher mismatch rate due to their texture richness. In contrast, images in flight strip 114-102, covering farmland with weaker textures, showed a lower mismatch rate. These results demonstrate the robustness of the OFFDIM method across various types of terrain and textures.

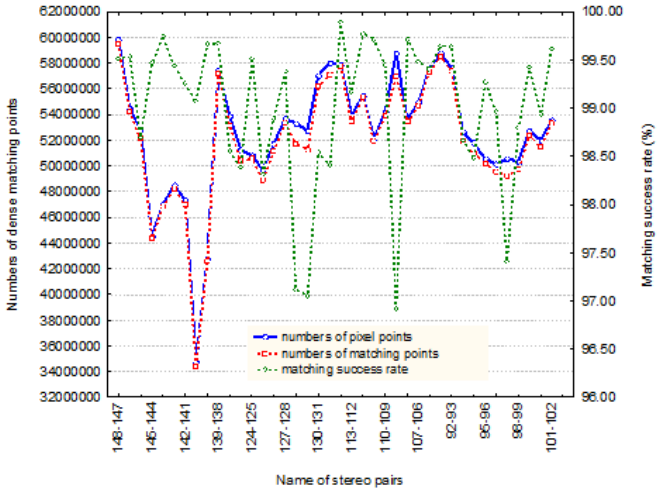


Figure 5. The analysis of Matching success rate (Yuan, 2018)

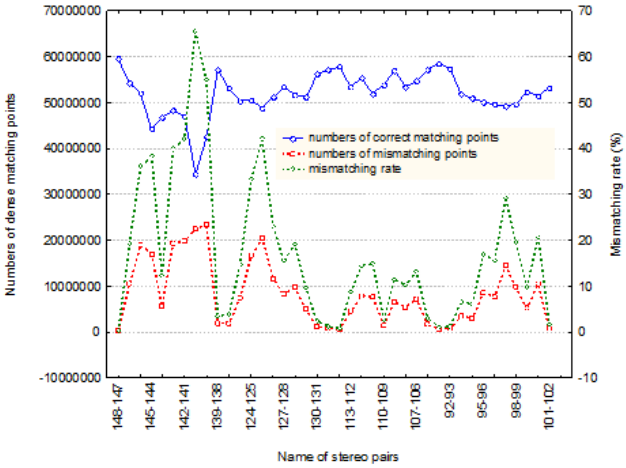


Figure 6. Mis-matching rate analysis (Yuan, 2018)

Figure 7 illustrates the variation in both image-side and in-ground accuracy for the pixel-wised dense image matching results within overlap areas of the stereoscopic image pairs. The figure clearly shows that the generated point clouds from OFFDIM achieve subpixel level on the image side, with all 40 stereoscopic image pairs exhibiting a matching accuracy greater than 1 pixel, and the highest reaching half pixel level. Consequently, the in-ground elevation accuracy of the automatically generated DTM is generally better than 20 centimeters, which is better than 3 ground sampling distance (GSD). However, there are variations in elevation accuracy between different models, ranging from 2.0 GSD to 3.5 GSD. A detailed comparison of the image-side and in-ground accuracy curves shows a completely consistent pattern of variation, further validating the robustness of the OFFDIM method.

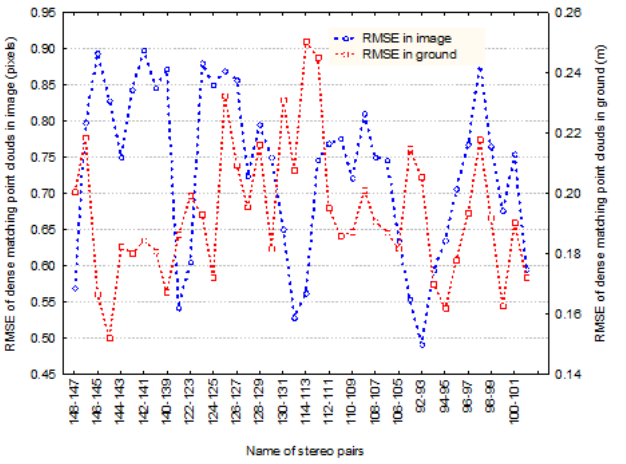


Figure 7. Matching accuracy analysis (Yuan, 2018)

4.4 Matching Efficiency Analysis

The experiments conducted in this paper were implemented on a desktop computer with an Intel(R) Core™ i7-6700HQ CPU, and 16.0 GB RAM. Figure 8 presents the CPU time consumption curve for dense image matching across 40 stereoscopic image pairs. According to the CPU time statistics, for UAV low-altitude photogrammetry images with a resolution of 11608 × 8708 pixels and a 60% forward overlap, the computational time for one stereo pair processing is from 192.593 to 327.246 seconds, demonstrating considerable efficiency. As PMVS can only perform dense matching for the entire strip collectively, Table 2 provides a comparison of CPU time consumption for dense matching of images from four strips using OFFDIM and PMVS, respectively, on a per-strip basis.

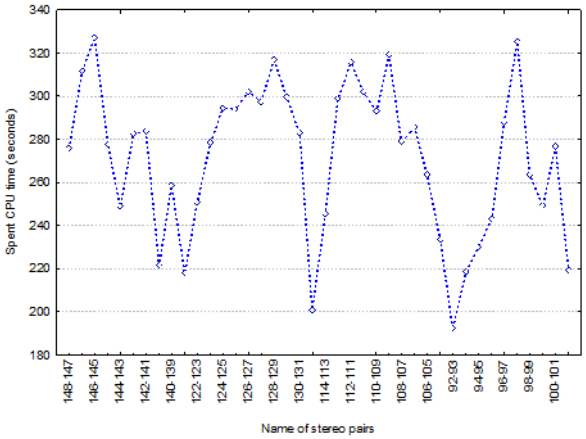


Figure 8. Computational Consumption time analysis (Yuan, 2018)

The computational efficiency shown in Table 2 demonstrated that the proposed method is approximately six times faster than that of traditional PMVS. Furthermore, a visual comparison of the 3D point clouds generated for the same area by PMVS and OFFDIM demonstrated that the OFFDIM-generated results are considerably higher than those of PMVS. When efficiency is evaluated based on the CPU time consumed per point, the proposed method exhibits superior matching efficiency compared to PMVS.

Strip No.	Number of Stereo Models	Computational Time (h: min : sec)	
		OFFDIM	PMVS
1	10	0:35:06	2:38:27
2	10	0:36:57	2:27:11
3	10	0:37:17	2:52:30
4	10	0:31:47	2:29:30

Table 2. The matching efficiency comparison (Yuan, 2018)

4.5 The Effect of Seed Points on OFFDIM

This paper conducts a quantitative analysis of the relationship between the number of seed points and the matching results in the OFFDIM method, which is designed to extract 3D point clouds from overlapping image regions using a dense matching approach based on a sparse optical flow field. The seed points utilized in this study originate from photogrammetrically densified points, exhibiting a generally uniform distribution within the overlap areas and achieving a measurement accuracy of better than ±0.15 pixels.

As illustrated in Figure 9, the accuracy of the dense matching point clouds on the image side improves as the number of seed points increases. However, once the number of seed points reaches approximately 1000, the accuracy stabilizes. Similarly, Figure 10 shows that the dense matching success rate increases with the number of seed points but also levels off around the 1000-point mark. The transition from sparse to dense optical flow requires fitting and interpolation, which means that the estimation time for the dense optical flow field increases as the

number of seed points grows. A comprehensive analysis of Figures 9 and 10 indicates that distributing around 1000 seed points uniformly within the overlap area is optimal for OFFDIM.

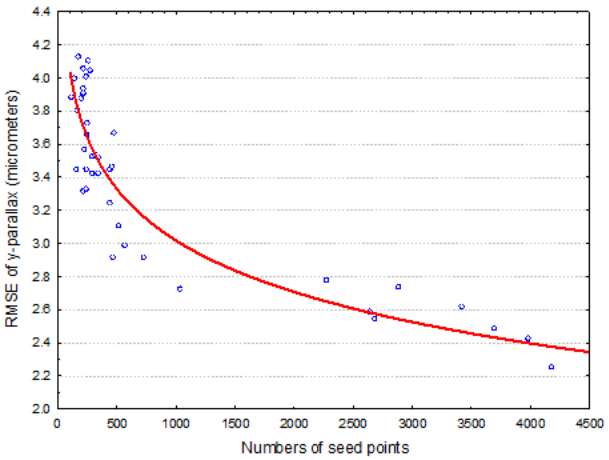


Figure 9. The relationship between number of seed points and output accuracy (Yuan, 2018)

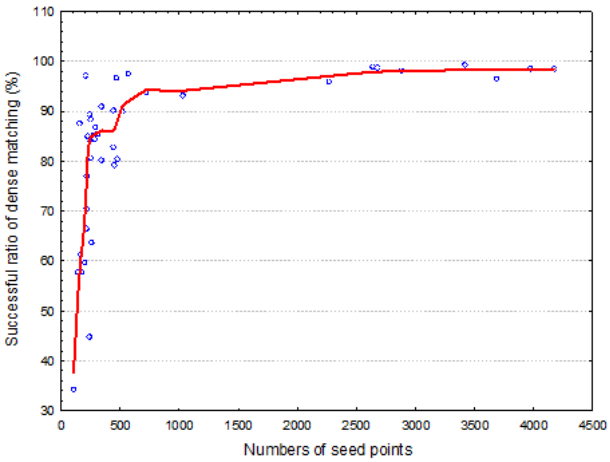


Figure 10. The relationship between the number of seed points and output reliability (Yuan, 2018)

5. Conclusion

Dense image matching is the bottleneck problem in photogrammetry and computer vision that requires urgent resolution. It has long been anticipated that this technology could rapidly transition from a research focus to a practical application, accelerating the automation of processes such as 3D reconstruction, Digital Elevation Model (DEM) extraction, and oblique photogrammetry for the retrieval of 3D geospatial information from images. This paper enhances a method that uses high-accuracy seed points as the initial input data to extract the dense optical flow fields in stereoscopic image pairs and refines matching points using epipolar constraints, resulting in an accurate and fast OFFDIM method capable of true pixel-level dense matching.

Extensive experiments conducted on low-altitude UAV photogrammetry images comprehensively evaluated OFFDIM’s performance in terms of both visual effects and in-ground quantitative metrics. The results demonstrate that OFFDIM achieves a matching success rate above 97% with sub-pixel matching accuracy, while the elevation accuracy of the automatically generated DTM is better than ±2 GSD. Additionally, OFFDIM proved to be more than five times more efficient than PMVS, offering higher matching success rates in areas with buildings and poor textures, and producing more complete dense point clouds that accurately represent terrain features.

The performance of the proposed method has a close connection to the distribution, accuracy, and number of the input seed points. In regions with a sufficient quantity of uniformly distributed seed points, dense matching results show significant improvement. Currently, the proposed methods operate solely with CPU, indicating potential for further efficiency enhancements. Future research will aim to explore GPU-based parallel algorithms. Reducing the algorithm's reliance on seed points and developing an effective seed point extraction model are essential.

Acknowledgments

This work is supported by the JSPS Kakenhi (Grants-in-Aid for Scientific Research, 21H05001, 23K13419) and the Cross-Ministerial Strategic Innovation Promotion Program (Grant Number JPJ012289)

References

Ai M., Hu Q., Li J., 2015. A robust photogrammetric processing method of low-altitude UAV images. *Remote Sensing*, 7(3), 2302－2333.

Chang J., Chen Y., 2018. Pyramid stereo matching network. In *IEEE Conference on Computer Vision and Pattern Recognition*, 5410– 5418.

Chapelle O. and Wu M., 2010. Gradient descent optimization of smoothed information retrieval metrics. *Inf. Retr.*, 13(3):216–235.

Chen Z., Sun X., Wang L., Yu Y., and Huang C., 2015. A deep visual correspondence embedding model for stereo matching costs. In *International Conference on Computer Vision*, 972–980.

Cheng X., Wang P., Yang R., 2018. Depth estimation via affinity learned with convolutional spatial propagation network. In *European Conference on Computer Vision*, 108–125.

Furukawa Y., Ponce J., 2010. Accurate, dense, and robust multi-view stereopsis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(98), 1362－1376.

Geiger A., Lenz P., and Urtasun R., 2012. Are we ready for autonomous driving? the KITTI vision benchmark suite. In *IEEE Conference on Computer Vision and Pattern Recognition*, 3354–3361.

Han X., Leung T., Jia Y., SukthankaR. R, Berg A. C., 2015. Matchnet: Unifying feature and metric learning for patchbased matching. In *IEEE Conference on Computer Vision and Pattern Recognition*, 3279–3286.

He K., Sun J., Tang X., 2013. Guided image filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(6), 1397-1409.

Hirschmuller H., Scharstein D., 2009. Evaluation of stereo matching costs on images with radiometric differences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(9), 1582－1599.

Issac H., Boykov Y., 2014. Energy based multi-model fitting and matching for 3D reconstruction. In *IEEE Conference on Computer Vision and Pattern Recognition*, 1146－1153.

Ke Y., Sukthankar R., 2004. PCA-SIFT:A more distinctive representation for local image descriptors. In *IEEE Conference on Computer Vision and Pattern Recognition,* 2, 506－513.

Kong D., Hai T., 2004. A method for learning matching errors for stereo computation. *BMVC*. 1, 2.

Remondino F., Spera MG., Nocerino E., et al., 2014. State of the art in high density image matching. *The Photogrammetric Record*, 29(146), 144－166.

Rock G., Ries J., Udelhoven T., 2011 Sensitivity analysis of UAV-photogrammetry for creating digital elevation models (DEM). *International Archives of Photogrammetry, Remote Sensing and Spatial Information Science*, 38(1), 1－5.

Rothermel M., Haala N., 2011. Potential of dense matching for the generation of high quality digital elevation models. In *ISPRS Workshop High-Resoultion Earth Imaging for Geospatial Information*, Hannover, Germany,38(4), 271－276.

Scharstein D., Szeliski R., Zabih R., 2001 A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. In *IEEE Workshop on Stereo and Multi-Baseline Vision*, 131－140.

Shao Z., Yang N., Xiao X., et al., 2016 A multi-view dense point cloud generation algorithm based on low-altitude remote sensing images. *Remote Sensing*, 8(5), 381.

Spyropoulos A., Komodakis N., Mordohai P., 2014. Learning to detect ground control points for improving the accuracy of stereo matching. *IEEE Conference on Computer Vision and Pattern Recognition*, 1621–1628.

Szeliski R., 2010, *Computer Vision: Algorithms and Applications*. Springer: Berlin, Germany.

Torresani L., Kolmogorov V., Rother C., 2013. A dual decomposition approach to feature correspondence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(2), 259－271.

Tola E., Lepetit V., Fua P., 2008 A fast local descriptor for dense matching. In *IEEE Conference on Computer Vision and Pattern Recognition*, 1－8.

Tran S., Davis L., 2006 3D surface reconstruction using graph cuts with surface constraints. In *European Conference on Computer Vision*, Graz, Austria, 219－231.

Turner D., Lucieer A., Watson C., 2012 An automated technique for generating georectified mosaics from ultra-high resolution unmanned aerial vehicle (UAV) imagery, based on structure from motion (SfM) point clouds. *Remote Sensing*, 4(5), 1392－1410.

Xiao X., Guo B., Li R., 2016 Multi-view stereo matching based on self-adaptive patch and image grouping for multiple unmanned aerial vehicle imagery. *Remote Sensing*, 8(2) 89.

Xu, D.; Ouyang, W.; Ricci, E.; Wang, X.; and Sebe, N. 2017. Learning cross-modal deep representations for robust pedestrian detection. In *IEEE Conf. on Computer Vision and Pattern Recognition*, 5363-5371.

Yuan X., Yuan W., Xu S., Ji Y., 2019. Research developments and prospects on dense image matching in photogrammetry. *Acta Geodaetica et Cartographica Sinica*, 48(12), 1542-1550.

Yuan X., 2008. A novel method of systematic error compensation for a position and orientation system. *Progress in Natural Science*, 18(8), 953－963.

Yuan, W., Yuan X., Cai Y., Shibasaki R., 2023. Fully automatic DOM generation method based on optical flow field dense image matching. *Geo-spatial Information Science*, 26(2), 242-256.

Yuan W., Yuan X., Xu S., Gong J., Shibasaki R., 2019. Dense Image-Matching via Optical Flow Field Estimation and Fast-Guided Filter Refinement. *Remote Sensing*, 11(20), 2410.

Yuan W., 2018. *An Automatic Digital Object Model Reconstruction from Optical Flow Field based Dense Aerial Image Matching*. Diss. The University of Tokyo.

Yuan W., Chen S., Zhang Y, et al., 2016 An aerial-image dense matching approach based on optical flow field. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 41(3), 543－548.

Zbontar J. and LeCun Y., 2015. Computing the stereo matching cost with a convolutional neural network. In *IEEE Conference on Computer Vision and Pattern Recognition*, 1592–1599.

Zhang L., Seitz S. M., 2007. Estimating optimal parameters for MRF stereo from a single image pair. *IEEE Trans. Pattern Anal. Mach. Intell*., 29(2), 331–342.

Zhou C., Zhang H., Shen X., 2017. Unsupervised learning of stereo matching. In *IEEE International Conference on Computer Vision*, 1567-1575.