PRECISE DISPARITY ESTIMATION FOR NARROW BASELINE STEREO BASED ON MULTISCALE SUPERPIXELS AND PHASE CORRELATION

Z. Ye[1], Y. Xu[1], L. Hoegner[1], X. Tong[2], U. Stilla[1]

[1]Photogrammetry and Remote Sensing, Technische Universität München, 80333 Munich, Germany
- (z.ye, yusheng.xu, ludwig.hoegner, stilla)@tum.de
[2]College of Surveying and Geo-Informatics, Tongji University, Shanghai 200092, China - xhtong@tongji.edu.cn

Commission II, WG II/2

KEY WORDS: Dense matching, Narrow baseline stereo, Subpixel phase correlation, Superpixel

ABSTRACT:

With the rapid development of subpixel matching algorithms, the estimation of image shifts with an accuracy of higher than 0.05 pixels is achieved, which makes the narrow baseline stereovision possible. Based on the subpixel matching algorithm using the robust phase correlation (PC), in this work, we present a novel hierarchical and adaptive disparity estimation scheme for narrow baseline stereo, which consists of three main steps: image coregistration, pixel-level disparity estimation, and subpixel refinement. The Fourier-Mellin transform with subpixel PC is used to co-register two input images. Then, the pixel-level disparities are estimated in an iterative manner, which is achieved through multiscale superpixels. The pixel-level PC is performed with the window sizes and locations adaptively determined according to superpixels, with the disparity values calcualted. Fast weighted median filtering based on edge-aware filter is adopted to refine the disparity results. At last, the accurate disparities are calculated via a robust subpixel PC method. The combination of multiscale superpixel hierarchy, adaptive determination of the window size and location of correlation, fast weighted median filtering and subpixel PC make the proposed scheme be able to overcome the issues of either low-texture areas or fattening effect. Experimental results on a pair of UAV images and the comparison with the fixed-window PC methods, the iterative scheme with fixed variation strategy, and a sophisticated implementation using global optimization demonstrate the superiority and reliability of the proposed scheme.

1. INTRODUCTION

Recovering the depth from stereo imagery is one of the cricual problems in photogrammetry. In conventional earth observation systems, for estimating elevation of the ground surface, one or more pairs of stereo images acquired by satellites or aircraft with a wide photogrammetric baseline are utilized, and the base to height (B/H) ratio of these stereo pair ranges from 0.6-1.0 (Morgan et al., 2010). Theoretically, a large B/H ratio is required, ensuring the accuracy of forward intersection for the elevation estimation. However, for the pair of stereo images with a wide baseline, it means that all the two images are acquired with totally different viewing angles. In such a situation, during the imaging process, 3D objects are recorded on the 2D image plane with different projection directions, which will generate different 2D patterns on the image for the same 3D object. This will hence increase more difficulties when identifying corresponding pixels in the image matching process (Delon , Rougé, 2007). Moreover, in the urban area, tall man-made infrastructures (e.g., skyscrapers or TV tower) will occlude lower neighboring objects (Xu et al., 2013), which will generate occlusions and shadows in the stereo images, making the matching of images more difficult.

To tackle those problems, the stereovision constructed by a narrow baseline could be one of the alternatives (Delon , Rougé, 2007). Precise and robust disparity estimation is highly demanded for narrow baseline stereo as the disparity precision greatly affects the height estimation. Fortunately, the developments of subpixel matching algorithm have enabled the estimation of image shifts with an accuracy higher than 0.05 pixel (Sabater et al., 2011, Tong et al., 2015), which makes the narrow baseline stereovision feasible. On the other hand, several challenges should be addressed to make disparity estimation accurate enough. The low-texture areas provide less information for matching and make the subpixel estimation unreliable especially in the case of small window size. In addition, when the correlation window strides across the depth discontinuities, the matching process suffers from the fattening effect that object boundaries are not reconstructed correctly. In this case, in addition to the subpixel matching algorithm, an effective matching scheme is also indispensable for narrow baseline stereo. In (Morgan et al., 2010), narrow baseline stereo matching was performed using a robust PC method with a fixed-scale matching window. In the work of (Takita et al., 2004, Arai , Iwasaki, 2012), a coarse to fine strategy based on image pyramid is adopted to improve the matching accuracy from the integer level to a subpixel level. While in the work of (Li et al., 2016), a hierarchical and adaptive framework is developed for the disparity estimation of UAV images, with a fixed variation strategy of window sizes and step size.

Inspired by these ideas, in this study we proposed a novel hierarchical and adaptive scheme for precise disparity estimation. The proposed scheme is based on multiscale superpixels and PC, with which we can reduce the influence of low-texture areas and fattening effect. Here, the challenge of low-texture areas is solved by a multiple-window strategy through multiscale superpixels in a hierarchical structure, with their reliability checked. While the fattening effect is addressed by adaptive determination of window size from the shape of superpixel, and using weighted median filtering based on edge-aware filter. The remainder of this paper is organized as follows. Section 2 provide a detailed explanation of our

proposed method. Section 3 presents the experiments and give a discussion and analysis of the derived results. Section 4 concludes the paper and plans future work.

2. METHODOLOGY

The implementation of the proposed hierarchical and adaptive disparity estimation scheme consists of three main steps: image coregistration, pixel-level disparity estimation, and subpixel refinement. In the first step, Fourier-Mellin transform with subpixel PC is used to obtain the global similarity transform model between two input images. Then, the pixel-level disparities are estimated in an iterative manner, which is achieved through multiscale superpixels. In each iteration, simple linear iterative clustering (SLIC) method (Achanta et al., 2012) is adopted to segment the input image into superpixels of different numbers. The pixel-level PC is performed with the window sizes and locations determined according to superpixels. A reliability check is implemented to ensure the robustness of low-texture areas. Subsequently, the pixels with the same superpixel label are filled with the same disparity values, and a shifting strategy updates the disparities. Finally, fast weighted median filtering based on an edge-aware filter is adopted to refine the disparity results. In the last step, the accurate disparities are calculated via the subpixel PC method using singular value decomposition and unified random sample consensus (SVD-RANSAC) (Tong et al., 2015). The overall workflow is illustrated in Fig. 1, and the detailed explanation on each step will be introduced in the following subsections.

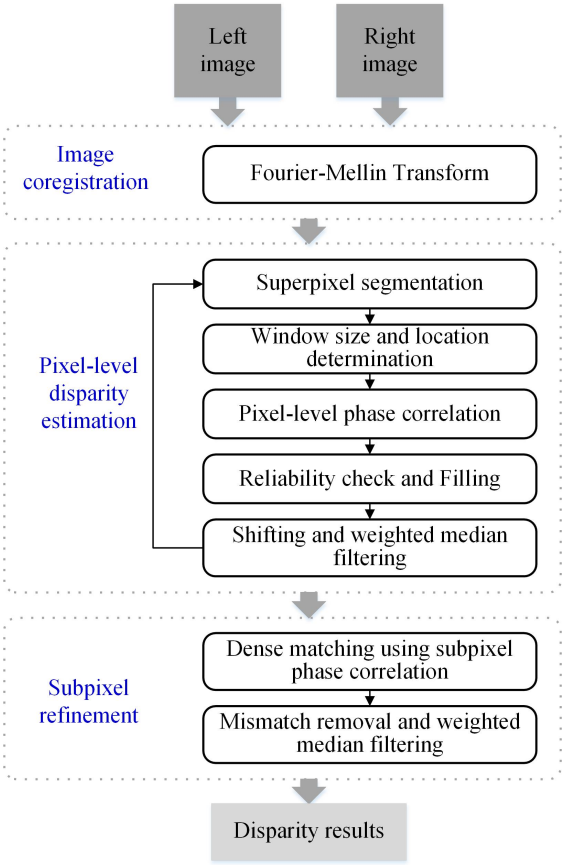

Figure 1. The workflow of the proposed hierarchical and adaptive disparity estimation scheme

2.1 Image coregistration with Fourier-Mellin transform

In order to achieve precise disparity estimation, the input images should be first aligned to eliminate the inconsistencies in addition to the disparity information. As the proposed scheme can directly estimate 2D displacement, the epipolar constraints is not strictly required. Feature-based or area-based registration methods (Zitova , Flusser, 2003) can alternatively be performed to reduce the search range and perspective distortions. In this study, as the narrow baseline stereo images have the similar viewing angle, the image registration method with Fourier-Mellin transform (Reddy , Chatterji, 1996) is employed to globally coregister the right image frame to the left via a similarity transform model.

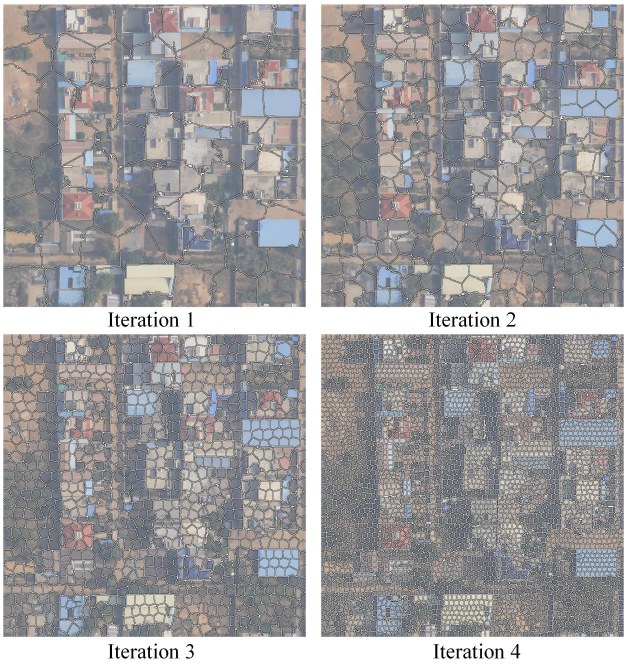

Figure 2. Multiscale SLIC superpixels in different iterations.

Image registration with Fourier-Mellin transform can account for translation, rotation and scale change between the images if these exist, and can be applied in the case of large motions without prior knowledge. Through the Fourier-Mellin transform, which corresponds to the log-polar mapping of the spectral magnitude, the rotation and scaling estimation can be represented as the translation estimation in an equivalent coordinate system. For accurate translation estimation, the SVD-RANSAC subpixel PC method is adopted.

2.2 Pixel-level disparity estimation

A novel hierarchical and adaptive framework is proposed to reduce the influence of low-textured areas and border regions. Different from the conventional manner that using image pyramid (Takita et al., 2004) or using fixed step size and window variation strategy (Li et al., 2016), the multiresolution and multiple-window PC is achieved via the multiscale superpixel segmentation. In each iteration, SLIC segmentation, window size, and location determination, pixel-level PC, reliability check, filling, shifting and fast weighted median filtering are implemented in order. In the next iteration, the

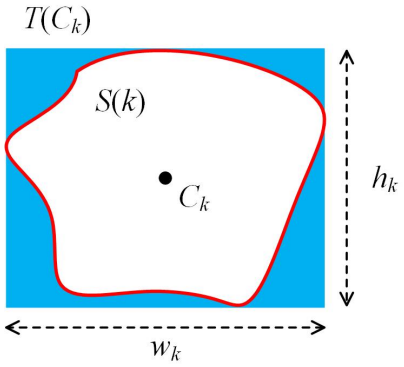

Figure 3. Determination of the window size and location of correlation.

number of superpixels is set to an increasing value to make the window size and step size of correlation gradually smaller, until it is up to a specified maximum number of iterations. We implement four iterations in this study.

The superpixel segmentation is crucial to the hierarchical and adaptive framework as we assume that the scene is piecewise continuous. SLIC superpixel method is selected due to its computational efficiency and excellent boundary adherence. Any other content-sensitive superpixel methods can be used. SLIC is regarded as an adaptation of k-means clustering to segmentation, in which a weighted distance measure combining color and spatial proximity is considered. As shown in Fig. 2, multiscale SLIC method decomposes the image into an increasing number of superpixels in the larger iteration. Each superpixel represents the corresponding object and adheres well to object boundaries. According to the shape of the superpixel segment, the window size and location of image correlation can be adaptively determined as illustrated in Fig. 3. With regard to each superpixel segment $S(k)$, the image correlation is carried out with the template window $T(C_k)$ which is the minimum bounding box centered at C_k. This adaptive determination strategy is able to minimize the influence of boundary overreach and fattening effect.

Phase correlation is adopted as the basic matching method in this study, which is a Fourier-based matching technique and is considered to be more accurate and effective than the commonly used area-based methods such as normalized cross-correlation (Ye et al., 2019). Phase correlation is based on the well-known Fourier shift property, which states that a shift of two relevant images in the spatial domain is transformed into the Fourier domain as linear phase differences. For each template window $T(C_k)$, assuming that two image functions $g_1(x, y)$ and $g_2(x, y)$ that are related by shifts x_0 and y_0 such that $g_2(x, y) = g_1(x-x_0, y-y_0)$. The normalized cross-power spectrum is given by:

$$Q(u, v) = \frac{G_1(u, v)G_2(u, v)^*}{|G_1(u, v)G_2(u, v)^*|} = \exp\{i(ux_0 + vy_0)\} \quad (1)$$

where $G_1(u, v)$ and $G_2(u, v)$ are the corresponding Fourier transform of $g_1(x, y)$ and $g_2(x, y)$, and $*$ denotes the complex conjugate. In the case of integer pixel shifts, the inverse discrete Fourier transform of $Q(u, v)$ is a unit impulse function centered on (x_0, y_0). Therefore, the pixel-level PC is realized by finding the peak coordinates of the inverse discrete Fourier transform of the normalized cross-power spectrum matrix:

$$(x_0, y_0) = \arg\max_{x,y} F^{-1}\{Q\}(x, y) \quad (2)$$

where F^{-1} denotes the inverse discrete Fourier transform.

The low-texture areas and dynamic-variation areas could significantly deteriorate the results of pixel-level PC in the case of smaller window size. Therefore, a reliability check similar to Li et al. (2016) is adopted before the shifting and updating operation. A decision threshold is adaptively estimated from the maximum peak correlation value of all the corrections in the first iteration. The reliability of each disparity from pixel-level PC is evaluated by comparing the peak correlation value with the decision threshold. For the reliable correlation, the pixels with the same superpixel label are filled with the same disparity values, and a shifting strategy updates the prior disparity map estimated in the previous iteration using the filled disparity map. For the unreliable correlation, the prior disparity map stops updating.

Due to the stepwise problem caused by superpixel filling and the presence of unreliable measurements, a constant time weighted median filtering method (Ma et al., 2013) is integrated for disparity refinement. The weights for median filtering are constructed using the constant time edge-aware filter, such as guided filter (He et al., 2013), which reduces the computational time and respects boundary structures. The weighted median filtering not only removes outlier errors but weakens the influence of the fattening effect. A disparity map is obtained after the weighted median filtering and is propagated to the next iteration.

2.3 Subpixel refinement

On the basis of the high-quality pixel-level disparity map generated from the previous step, the subpixel accuracy is pursued in this step by means of the subpixel PC. The window size for subpixel PC can be determined according to the results of reliability check in different iterations in the pixel-level disparity estimation step. The smaller window size is adequate if the reliability checks in all iterations pass.

The SVD-RANSAC subpixel PC method (Tong et al., 2015) is adopted due to its high reliability and strong robustness. The SVD-RANSAC method integrates the advantages of using SVD algorithm to converts the translation estimation problem to one dimension and using RANSAC algorithm for robust linear fitting. After calculating the normalized cross-power spectrum matrix as in Equation (1), we extract the phase difference in each dimension from the masked and filtered normalized cross-power spectrum matrix using SVD and 1-D unwrapping. The slopes of the unwrapped phase angles of the left and right dominant singular vectors are estimated using the RANSAC algorithm and converted to the subpixel shifts.

Similar to the pixel-level disparity estimation, the subpixel disparity map is further refined through mismatch removal based on the matching uncertainty measures outputted from the SVD-RANSAC method as well as weighted median filtering.

3. EXPERIMENTS AND RESULTS

3.1 Experimental data

In order to demonstrate the performance of the proposed method, a subset pair of successive images with a B/H ratio of

Figure 4. UAV test images and image coregistration.

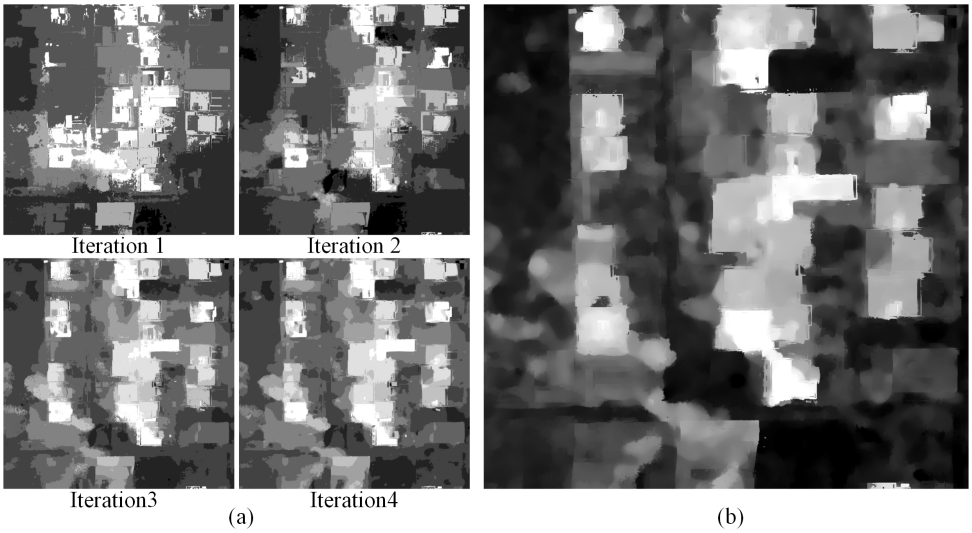

Figure 5. Results of disparity estimation in the (a) pixel-level steps and (b) subpixel refinement step of the proposed scheme.

lower than 0.1 is adopted as shown in Fig. 4. The images were captured by unmanned aerial vehicle (UAV) with a NIKON D800 digital single-lens reflex camera. The imaging scene is an urban area with a number of buildings, which is suitable to investigate the influence of fattening effect. Image registration with Fourier-Mellin transform was carried out to estimate the global similarity transform between two images, and the right image was accordingly warped. After image coregistration, the parallax disparities resulted from the relative height variation mainly lie in the y-direction. Therefore, only the y-direction disparity maps are displayed.

3.2 Evaluation of subpixel matching accuracy

To evaluate the subpixel matching accuracy of the SVD-RANSAC algorithm we used, we conducted experiments using simulated data generated from the UAV images, with the approach given in (Tong et al., 2015). The baseline PC based methods we compared include Stone's (Stone et al., 2001), Leprince's (Leprince et al., 2007), and PEF (Nagashima et al., 2006). In these experiments, two dominant error sources of corruptions, affecting the performance of PC are analyzed. The first experiment is the Aliasing experiment, referring to an effect that results in different signals becoming indistinguishable during sampling. While the second one is the Noise experiment, which is to test the robustness of methods when data is contaminated by noise coming from both the ground disturbance and optical systems.

For the Aliasing experiment, the corresponding results of various PC methods in terms of the mean value (MV) and the root mean square errors (RMSE) as a function of σ are shown in Fig. 6. Here, σ denotes the standard deviation of a Gaussian filter that controls the amount of aliasing. While for the Noise experiment, the corresponding results of various PC methods regarding the MV and the RMSE as a function of V_n are given in Fig. 7. Here, V_n stands for the normalized variance of noise added. As seen from the figure, it is clear that in the presence of both aliasing and noise, the SVD-RANSAC method can always outperform other PC methods, providing a similar but better performance as the Leprince's method. To be specific, in the result of Aliasing, the MV and RMSE are always less than 0.02 pixels. Whereas in the result of Noise, the MV and RMSE can be less than 0.2 pixels even with the V_n reaching 0.04.

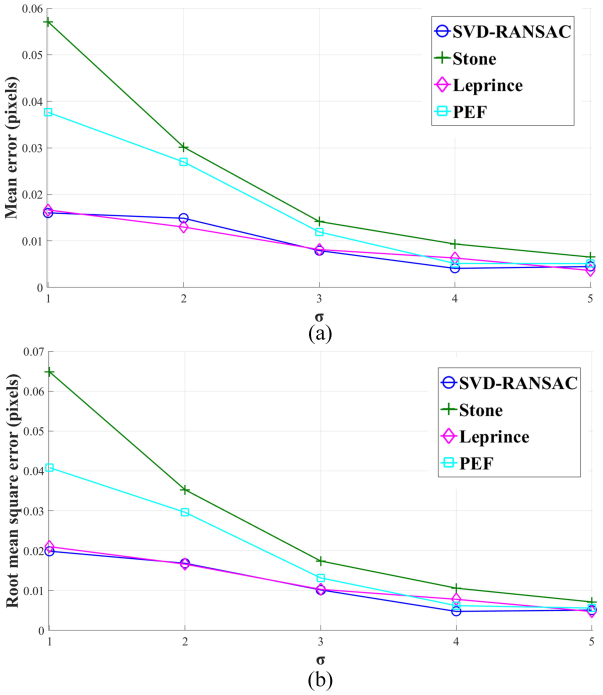

Figure 6. Results of the Aliasing experiment: (a) MV and (b) RMSE.

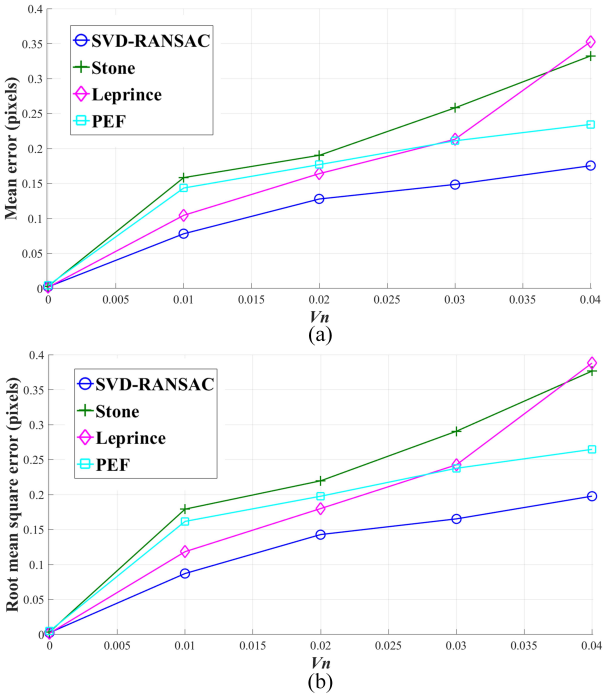

Figure 7. Results of the Noise experiment: (a) MV and (b) RMSE.

3.3 Results of disparity estimation

Fig. 5 plots the results of the proposed disparity estimation scheme based on multiscale superpixel and PC in the pixel-level step and subpixel refinement step. It can be found from the coarse-to-fine process in Fig. 5(a) that the details of the disparity map are gradually recovered and the messy areas are consistently decreasing with the rising number of superpixels in the successive iterations. This confirms the feasibility of our hierarchical and adaptive framework constructed via multiscale superpixel and fast weighted median filtering. The disparity map generated from the previous iteration propagates good initial values to the next iteration and finally provides a satisfied condition for the subpixel refinement. Visual investigation of Fig. 5(b) shows that the final disparity map after the subpixel refinement mainly reflects the height difference between the buildings, and the depth discontinues accord well with the object boundaries.

3.4 Comparison with other implementations

To further evaluate the performance of the proposed disparity estimation scheme, we compare the disparity map with the ones from three other dense matching implementations, including fixed-window pixel-by-pixel PC, a hierarchical and adaptive framework with fixed variation strategy (Li et al., 2016) and MicMac (Pierrot-Deseilligny , Paparoditis, 2006). For fixed-window PC method, the popular implementation COSI-Corr (Leprince et al., 2007) is adopted. Two window sizes, i.e., a larger value of 64*64 and a smaller value of 16*16, are tested. Unweighted median filtering is additionally applied to smooth the disparity map as suggested in (Morgan et al., 2010). For all the implementations, we register the input images in advance using Fourier-Mellin transform.

Fig. 8 displays the results of the disparity map generated by the proposed and other dense matching schemes. As can be seen from the results of fixed-window COSI-Corr in Fig. 8 (a) and (b), the disparity map suffers from the influence of either low-texture areas or boundary overreach. In the case of larger window size, the fattening effect is serious as the correlation window is likely to stride across the depth discontinuities. In contrast, the fattening effect is weakened in the case of smaller window size. However, the disparity map is relatively noisy as the correlation window provides inadequate information in the low-texture areas. The hierarchical and adaptive framework proposed in (Li et al., 2016) can reduce the influence of the low-texture regions and boundary overreach to some extent, but the performance is still unsatisfied in the presence of complicated depth discontinuities such as in the urban areas, since the fixed variation strategy of window size and step size is adopted. In addition, it can be inferred from the above three cases that the unweighted median filtering has little effect on solving the fattening effect, although it can filter the apparent mismatches. MicMac finds the disparity map that minimizes an energy function using the multi-directional dynamic programming associated with a multiresolution strategy. A regularization term is embedded in considering the neighbor information. Therefore, it provides good results when using small correlation windows. The proposed scheme achieves the similar performance with MicMac, which valid the ability to deal with the issues of both low-texture areas and fattening effect. On the other hand, the proposed scheme using local matching method possesses the potential advantage of higher computational efficiency compared to the time-consuming global optimization matching methods. Besides, we examine the disparity of five disparity estimation schemes along a random profile (see Fig. 9). The curves of the proposed scheme and MicMac are the most similar ones with each other, which confirms the relative accuracy of the proposed disparity estimation scheme.

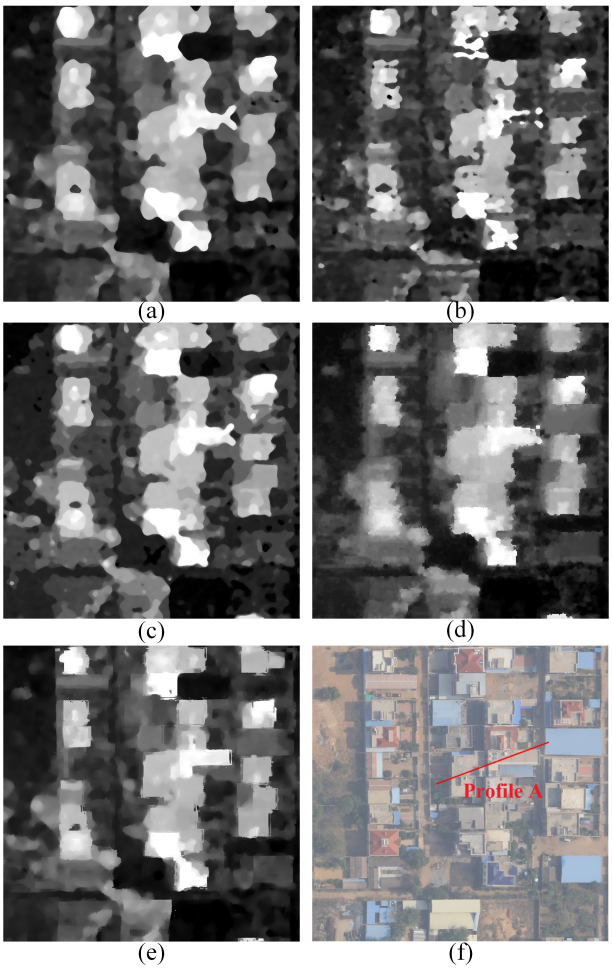

Figure 8. Disparity map generated by (a) COSI-Corr with a larger window size, (b) COSI-Corr with a smaller window size, (c) hierarchical and adaptive framework with fixed variation (HAFV), (d) MicMac and (e) the proposed scheme. (f) Original left image.

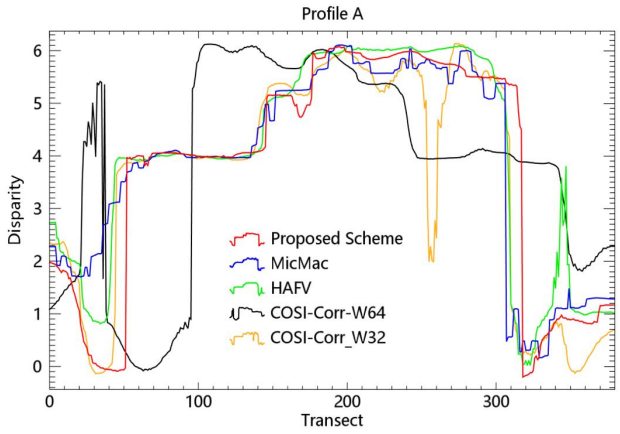

Figure 9. Comparison of the disparity along a random profile (see Fig. 8f).

4. CONCLUSION

In this study, we introduce a novel hierarchical and adaptive disparity estimation scheme for narrow baseline stereo. The integration of multiscale superpixel hierarchy, adaptive determination of the window size and location of correlation, fast weighted median filtering and subpixel PC make the proposed scheme be able to overcome the issues of both low-texture areas and fattening effect. The experimental results of our scheme on a pair of UAV images outperform those of the fixed-window PC methods, the iterative scheme with fixed variation strategy, and are on par with some sophisticated implementations using global optimization, such as MicMac. In future work, shadow detection and construction of weights with other factors will be considered to reduce the sensitivity to the unexpected image intensity.

ACKNOWLEDGEMENTS

The authors would like to appreciate Dr. Xiangfeng Liu from Shanghai Institute of Technical Physics, Chinese Academy of Sciences for providing the UAV test data.

REFERENCES

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Ssstrunk, S., 2012. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34, 2274-2282.

Arai, T., Iwasaki, A., 2012. Fine image matching for narrow baseline stereovision. *IEEE International Geoscience and Remote Sensing Symposium*, 2336–2339.

Delon, J., Rougé, B., 2007. Small baseline stereovision. *Journal of Mathematical Imaging and Vision*, 28, 209–223.

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

Leprince, S., Barbot, S., Ayoub, F., Avouac, J., 2007. Automatic and precise orthorectification, coregistration, and subpixel correlation of satellite images, application to ground deformation measurements. *IEEE Transactions on Geoscience and Remote Sensing*, 45, 1529-1558.

Li, J., Liu, Y., Du, S., Wu, P., Xu, Z., 2016. Hierarchical and adaptive phase correlation for precise disparity estimation of UAV images. *IEEE Transactions on Geoscience and Remote Sensing*, 54, 7092–7104.

Ma, Z., He, K., Wei, Y., Sun, J., Wu, E., 2013. Constant time weighted median filtering for stereo matching and beyond. *Proceedings of the IEEE International Conference on Computer Vision*, 49–56.

Morgan, G. L. K., Liu, J. G., Yan, H., 2010. Precise subpixel disparity measurement from very narrow baseline stereo. *IEEE Transactions on Geoscience and Remote Sensing*, 48, 3424-3433.

Nagashima, S., Aoki, T., Higuchi, T., Kobayashi, K., 2006. A subpixel image matching technique using phase-only correlation. *International Symposium on Intelligent Signal Processing and Communications*, 701–704.

Pierrot-Deseilligny, M., Paparoditis, N., 2006. A multiresolution and optimization-based image matching approach: An application to surface reconstruction

from SPOT5-HRS stereo imagery. *ISPRS Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36.

Reddy, B. S., Chatterji, B. N., 1996. An FFT-based technique for translation, rotation, and scale-invariant image registration. *IEEE Transactions on Image Processing*, 5, 1266-1271.

Sabater, Neus, Morel, J-M, Almansa, Andrés, 2011. How accurate can block matches be in stereo vision? *SIAM Journal on Imaging Sciences*, 4, 472–500.

Stone, Harold S, Orchard, Michael T, Chang, Ee-Chien, Martucci, Stephen A, 2001. A fast direct Fourier-based algorithm for subpixel registration of images. *IEEE Transactions on geoscience and remote sensing*, 39, 2235–2243.

Takita, K., Muquit, M. A., Aoki, T., Higuchi, T., 2004. A sub-pixel correspondence search technique for computer vision applications. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, 87, 1913–1923.

Tong, X., Ye, Z., Xu, Y., Liu, S., Li, L., Xie, H., Li, T., 2015. A novel subpixel phase correlation method using singular value decomposition and unified random sample consensus. *IEEE Transactions on Geoscience and Remote Sensing*, 53, 4143-4156.

Xu, Y., Ye, Z., Li, L., Liu, S., Li, T., Tong, X., 2013. Generating dem of very narrow baseline stereo using multispectral images. *International Conference on Remote Sensing, Environment and Transportation Engineering*, Atlantis Press.

Ye, Z., Tong, X., Zheng, S., Guo, C., Gao, S., Liu, S., Xu, X., Jin, Y., Xie, H., Liu, S., Chen, P., 2019. Illumination-robust subpixel Fourier-based image correlation methods based on phase congruency. *IEEE Transactions on Geoscience and Remote Sensing*, 57, 1995-2008.

Zitova, B., Flusser, J., 2003. Image registration methods: a survey. *Image and Vision Computing*, 21, 977–1000.

Revised April 2019