The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLIII-B3-2020, 2020 XXIV ISPRS Congress (2020 edition)

STACKED LOCAL FEATURE DETECTOR FOR HYPERSPECTRAL IMAGE

Zhao Yan^{1,*}, Zhaocong Wu¹

¹ School of Remote Sensing and Information Engineering, Wuhan University, Hubei, China - (whuhenry, zcwoo)@whu.edu.cn

Commission III, WG TCIII/3

KEY WORDS: Stacked Feature Point, hyperspectral image registration, keypoint detection

ABSTRACT:

Images registration is an important task in hyperspectral image processing, while almost all local feature point based image matching algorithm is designed for single band image only and miss the advantage of abundant spectral information. Therefor, in this paper we propose a novel local feature detector for hyperspectral image. This method, which is named stacked local feature detector (HSI-SFD), stack all local feature points detected from every single spectral band. With redundant information, local feature miss-detected at noise pixel can be filtered out by the stack count threshold, leading to more reliable and robust local features. To verify the algorithm, a hyperspectral image matching dataset is built. Multiview hyperspectral images of the several different flat targets are taken by scan-line hyperspectral camera in library. Each image contains 270 bands within 400-1000nm. Ground-truth transformation matrix between images is computed from corresponding points selected by hands. Experiment on the dataset shows that features points miss-detected at texture-less area can be inhibited by HSI-SFD. Both keypoint repeatability and matching accuracy increase significantly. Precision and matching score increase up to 10% for some scene.

1. INTRODUCTION

Due to the recent advances of hyperspectral sensors, application and research of hyperspectral images (HSI) extended from satellite based remote sensing to close range target detection and recognition, such as 3D forestry mapping, health analysis, food security, etc. Among these applications, images registration is a basic and important pre-processing step, while almost all the image registration algorithms can handle single band image only and miss the advantage of abundant spectral information. In addition to this, HSI registration suffer from image noise due to narrow band of light captured by sensor within in a short period of time. The image noise are usually local peaks and tend to be mis-detected as local feature by algorithm such as SIFT or ORB.

Therefore, in this paper we propose a novel local feature detector for hyperspectral images. By stacking local feature points on different spectral bands and filtering out ones that do not contains enough bands keypoints, stack local feature points is more reliable than any local features on any single band. The contribution of this paper are as follows:

- A new local feature detector designed for hyperspectral image is proposed. Spectral and spatial information are fully used to extract reliable keypoints by stacking local features from different bands. With redundant spectral information, the detector is rubost to random noise on single band. Besides, the algorithm is easy to be implemented in parallel, leading to much faster detecting process and less memory requirment than 3D-SIFT based detector.
- 2) A new hyperspectral image matching dataset is created. Multiview hyperspectral images of the 4 different flat target is taken by scan-line hyperspectral camera. We propose a geometrical model to describe the geometry relationship between scan-line image pair by re-projecting the original

images to an equivalent frame image. Ground-truth of transformation matrixes is also provided.

2. RELATED WORK

Local feature detector and descriptor is a fundamental, yet important task in computer vision and photogrammetry. The research history of this issue can be traced back up to 1980s. Early detectors were artificially designed to find corners or local peak points on both intensity and gradient, such as Harris, DoG and Hessian. These detectors are considered to be viewport invariant, but sensitive to scale change. With introduction of Scale Space Theory, improved methods, such as SIFT(Lowe, 2004) and SURF(Bay et al., 2008), are both scale and viewport invariant and applied successfully in image registration, 3D reconstruction, image retrieval and many other fields. In the past few years, deep learning has revolutionized computer vision, pushing the community to propose deep-learning methods for local feature detection and description. The problem of constructing deep-learning based local feature detector is much more complicated than other computer vision tasks, because there is no standard for what is local feature. Local feature identified by different person may vary dramatically. As a result, there is no ground-truth label and no neural network can be trained in a supervised way. Deep learning keypoint detector needs to be designed in a new formulation. First attempt was DetNet(Lenc, Vedaldi), in which learning task of keypoint detector is to find points that satisfy the so-called covariance constraint, rather than the labeled location. Other studies, such as TCovNet(Zhang et al., 2017) and D2-Net(Dusmanu et al., 2019) optimize the idea and get better result. Another attempt was SuperPoint(DeTone et al., 2018) and UnSuper-Point(Christiansen et al., 2019). They used the synthetic images with well accepted keypoint location, such as standalone points or corners, as ground truth labels. However, recent study(Jin et al., 2020; Csurka et al., 2018) indicates that deep-learning based method do not perform as well as they declared in real world

^{*} Corresponding author

dataset, and sometimes even worse than handcrafted local feature detector.

However, almost all keypoint detector are developed to work with single band image, no matter handcrafted or deep-learning based ones. As hyperspectral image is widely used in various domains, local feature detector designed for hyperspectral image draws more and more attention. Algorithms were developed to get more reliable and robust keypoints by taking advantage of spectral information available in multi and hyperspectral image. A natural solution is to generalize traditional 2D keypoint detector to \mathbb{R}^n spaces (especially \mathbb{R}^3 for hyperspectral datacube). This idea come from volumetric medical image registration, such as computed tomography (CT) and magnetic resonance imaging (MRI). 3D Sift(Rister et al., 2017), for example, use 3D DoG to locate keypoints and eigenvector of structure tensor to derive a local orientation to each keypoint. However, there is a dearth of research of applying 3D Sift to HSI data, because the third dimension of hyperspectral image, which is reflectance change across spectral wavelength, has different physical interpretation from medical images. As a result, SS-Sift (Al-khafaji et al., 2018)and UMSGC-SIFT (Li et al., 2018) was proposed to detect keypoints for HSI data. In SS-Sift, 3D Gaussian filte is calculated with different param in spatial and spectral axis, to make stronger smoothing effect in spectral domain and a better result for keypoint detection. In UMSGC-SIFT, both gradient and intensity value are considered to overcome the low signal-noise ratio. Both algorithms can make full use of all band information to locate the keypoints on either spectral domain or spatial domain. The main drawback of these detector is memory consuming. Due to calculation of 3D-DoG of whole HSI data cube, runtime memory requirement is several times, or even tens of times, the size of data cube. The spectral and spatial resolution of HSI data, as hyperspectral image sensor developed, have increased significantly. The size of single HSI data cube can reach even tens of GB with more than 200 bands and eight million pixels, which makes it is impossible to run such detector on the whole HSI image without cutting it apart. On the other side, hyperspectral images tend to include significant redundancy in adjacent bands, leading to large amounts of redundant information being processed, stored, and transmitted (Feng et al., 2017). Inspired by hyperspectral data classification, band selection (Ordóñez et al., 2018) or feature extraction(Mukherjee et al., 2009) of HSI image is processed to reduce spectral redundancy before local feature detection. Then local feature is detected on every selected or extracted bands, as a result, memory requirement is quite low and is irrelevant to spectral bands count. Meanwhile, detector of each band is independent, making it easy to be implemented in a parallel way(Ordóñez et al., 2020). However, keypoints obtained by these methods are simply a collection of keypoints of all bands, failing to improve keypoint repeatability and reliability by making use of spectral information across wavelength.

3. METHOD

3.1 Stacked Keypoint Detection

In order to facilitate the simultaneous use of HSI spectral and spatial information, we developed a framework called Hyperspectral Stacked Feature Detector (HSI-SFD). The framework can integrate feature points of all spectral bands. As shown in Figure 1, there are 3 main steps of feature detection and description in the proposed method: 1) Single Band Keypoint Detection: Feature point is detected by a certain feature point detector, such as SIFT, ORB or modern deep-learning based feature detector, for each band of hyperspectral image. Keypoint collection \mathbf{P}^k of kth band can be expressed as follows:

$$\mathbf{P}^k = Det(\mathbf{I}_B^k) \tag{1}$$

where \mathbf{I}_B^k is the *k*th band of the hyperspectral image **I**. In this paper, SIFT detector (implemented by VLFeat(Vedaldi, Fulkerson)) is used in this paper, for it is wildly tested and proven to be the golden standard in local feature detection, even compared with modern deep-learning based keypoint detector(Csurka et al., 2018)

2) Stack Keypoints by Location: It is hard to find keypoints at the same coordinate on different bands, because keypoints may hay sub-pixel coordinate and there may be slight difference for the same points on different bands. To simplify the problem, stack points are supposed to locate only on the center of each pixel, which means the coordinates of stack points are integers. The stack point located at pixel (m, n), $\mathbf{P}_{(m,n)}^S$, is generated from keypoints from every bands that meet the requirement as follows:

$$\mathbf{P}_{(m,n)}^{S} \Leftarrow \{\mathbf{P}_{(x,y)}^{k} \mid \sqrt{(m-x)^{2} + (n-y)^{2}} < \varepsilon_{s}\} \qquad (2)$$

where $\mathbf{P}_{(x,y)}^{k}$ is the keypoint located at (x, y) on the kth band. ε_s is stack threshold to determine the scope of one stackpoints. Stack points consisting of t feature points at different bands is called t-stacked feature points \mathbf{P}_t^S and is expressed as follows:

$$\mathbf{P}_{t}^{S} = \{ \mathbf{P}_{(m,n)}^{S} \mid ConsNum(\mathbf{P}_{(m,n)}^{S}) = t) \}$$
(3)

where $ConsNum(\mathbf{P}_{(m,n)}^S)$ is the count of feature points the stack point $\mathbf{P}_{(m,n)}^S$ consists of.

3) Keypoint Description By Panchromatic Image: Local features of corresponding panchromatic image are detected and described by the same detector in step 1). Then panchromatic keypoints within ε_p pixel from the nearest stacked feature point is retained. Their descriptor is used as the stacked feature point descriptor.

$$\mathbf{Desc}_t^s = \mathbf{Desc}_{pan} \quad where \ d(\mathbf{P}_{pan}, \ \mathbf{P}_t^S) < \varepsilon_p \qquad (4)$$

where \mathbf{Desc}_{pan} is the descriptor of panchromatic keypoint \mathbf{P}_{pan} . $d(\mathbf{P}_{pan}, \mathbf{P}_t^S)$ is the smallest Euclidean distance between \mathbf{P}_{pan} and t-Stack Keypoints Set \mathbf{P}_t^S

Since most hyperspectral camera can not capture corresponding panchromatic image, it is synthesized from hypercube in our research. 4 different decolorization algorithms are used to generate panchromatic image, which are 1, Reflectance Mean of all bands; 2, Radiance integration of all bands; 3, FakeGray: weighted sum of fake RGB bands (640nm, 550nm and 470nm); 4. Contrast Preserved Decolorization algorithm proposed by Kang et al. 2018.

3.2 Evaluation Metric

The performance of HSI-SFD is evaluated through image registration by KNN matching (k=2 in our work). Radio test is applied to eliminate false-positive matches and a threshold of 0.8 The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLIII-B3-2020, 2020 XXIV ISPRS Congress (2020 edition)



Figure 1. Schematic of the proposed stacked local feature detector for hyperspectral image



Figure 2. Pseudo-color images of the samples in the database

is used for all test cases. We use the metrics of putative match ratio, precision, matching score, recall, proposed by Heinly et al. 2012, to evaluate feature matching. Besides, we also compute the widely used standard metric: repeatability, to test the performance of detector only.

4. EXPERIMENT

There are two experiments taken in this paper. One studies the influence of stack count on keypoints count and distribution. Another one analyses the match performance of HSI-SFD. Details and datasets we used are described as follows.

4.1 Dataset

As there is no publicly available hyperspectral image matching dataset, a new affine HSI dataset was created to evaluate hyperspectral feature detector and descriptor. Inspired by widely used Oxford VGG-affine dataset and HPatches dataset, geometry relationship between hypercubes in the dataset can be well approximated by homography matrix. In detail, 8 sequence were captured by Headwall Nano-Hyperspec camera in laboratory. All targets are flat object made of different material. Viewport is translated horizontally in 4 sequence (translation dataset) and rotated in the other 4 scenes (rotation dataset). There are 6 images in translation dataset and 5 images in rotation dataset. Pseudo-color images of some samples in the dataset are show in Figure 2. Each hypercube contains 270 spectral bands within 400-1000nm. Only 163 bands in visible bands, whose wavelength range is 400-760nm, is used to synthesize panchromatic image. Stack points are also detected in these bands for consistency.

Pixel value of original hyperspectral images is digital number(DN), and it is influenced by both incident light intensity and material reflectivity. To get the reflectance data, a reflectance panel is used the estimate the incident light. The light source we used is a halogen lamp and its light field is uneven. To overcome this issue, we propose a preprocess shown in figure 3. Firstly, image of reflectance panel is captured. Then target is placed just in front of the panel and image is taken by the same parameters. Follow by dark current correction and back ground removal, radiance image for both target and panel are generated. Finally reflectance image is generate by division and all pixels which does not belong to the target are removed by masking and croping.

The Nano-Hyperspec is a line-scanning hyperspectral camera. As a result, geometry relationship between scan-line hyperspectral images cannot be modelled by affine transformation. In order to address this issue, our solution is to generate equivalent virtual frame hyperspectral image from the original one. Specifically, virtual image plane is defined the one same as central scanline. Then original images are re-projected to the equivalent image plane by the line-scan camera geometrical model presented by Ait-Aider Berry. Camera was calibrated using the Close-Range Photogrammetry Calibration Field of School of Remote Sensing and Information Engineering, Wuhan University. Finally, ground-truth affine transformation matrix is com-

The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLIII-B3-2020, 2020 XXIV ISPRS Congress (2020 edition)



Figure 3. Preprocess of hyperspectral data to generate reflectance image



Figure 4. Number of Stack Points with respect to Stack Count

puted from the corresponding points selected by hands.

4.2 Keypoint distribution

Stack points was detected under different stack count on images within one dataset and result of BookCover Lica dataset is shown in Figure 4. As we expected, the count of stacked feature points decreases as stack count increase. Besides, the way stack points count decrease shows similar patterns among images. From 1 to 10 stacked layers, stacked feature points numbers are rapidly diminishing. Most of filtered features are invalid ones which can hardly find any matches and are caused by image noise. Besides, kernel density analysis shows that features located at texture-less region reduce significantly, as shown in Figure 5. There is almost no keypoints in these areas when stack count reaches 10. The feature point matching also verified that there is no effective match among these features. Both results indicate that HSI-SFD can make use of band information to overcome low SNR and get more reliable features. When stacked band is larger than 10, stacked feature point decrease in proportion to stacked band count. For each addition stack count, the number of keypoints decreases by about 4%. The result indicates that no feature point can be detected in all spectral bands. With more spectral information, more local feature at different spectral range can be detected.

4.3 Keypoints Matching

First, keypoints on all 8 scenes is detected and described by Sift algorithm and matching between all image pairs using KNN algorithm. The results are listed in Table 1 & 2. It indicates that Radiance Integration reports the lowest value of all metrics and 10% approximately less than other decolorization method. The other 3 methods show similar match performance. However



(c) Pseudo Color

Figure 5. Kernel aensity analysis of 1 & 10 stack points. Blue means there is no keypoints while green represent keypoints. The brighter the color, the more the keypoints.

Table	1. Matching Metric for	panchromatic keypoints or	n
	translation dataset (all	data are percentage)	

	PMR	Precision	MS	Recall	Repeatability
AveRef	50.54	67.90	36.11	64.59	50.59
Contrast Preserve	46.14	65.85	31.63	60.04	48.20
Radiance	38.52	55.81	23.28	52.86	38.62
FakeGray	51.57	68.36	36.70	68.20	49.47

FakeGray is slightly better, about 1-3% higher in all metrics on all test cases.

Then, HSI-SFD is applied to all images and 10-stacked feature point is detected and matched between all image pairs. The result is shown in Table 3 & 4. All performance metrics increase on all test cases. For example, putative match ratio of the best performance test case, FakeGray, increase from 51.57% to 60.44%, precision increase from 68.36% to 72.71%, matching score increase from 36.7% to 44.60%, recall increase

Table 2. Matching Metric for panchromatic keypoints on translation dataset (all data are percentage)

	PMR	Precision	MS	Recall	Repeatability
AveRef	42.71	75.62	33.71	64.58	48.80
Contrast Preserve	42.12	76.05	32.51	63.64	48.77
Radiance	30.70	62.04	20.74	49.32	38.60
FakeGray	47.62	77.11	37.51	70.77	50.31

	PMR	Precision	MS	Recall	Repeatability
AveRef	60.44	72.27	44.59	74.19	56.38
Contrast Preserve	56.84	70.24	40.48	71.57	53.67
Radiance	49.81	69.35	35.17	67.15	48.69
FakeGray	60.17	72.71	44.60	73.83	56.78

 Table 3. Matching Metric for 10-stack points on translation

 dataset (all data are percentage)

 Table 4. Matching Metric for 10-stack points on rotation dataset

 (all data are percentage)

	PMR	Precision	MS	Recall	Repeatability
AveRef	51.50	79.84	41.61	69.51	57.63
Contrast Preserve	52.48	80.40	42.35	71.40	57.39
Radiance	42.85	77.88	33.73	61.74	51.91
FakeGray	55.57	81.80	45.71	75.16	58.60

from 68.20% to 74.19% and repeatability increase from 50.59% to 56.78%. Among all different decolorization methods, 10-stacked feature points of FakeGray and Reflectance Mean algorithm show much the same result on putative match radio, matching score, recall and repeatability, which are 4 5% better than the other two decolorization methods and 2% better on precision. For translation and rotation dataset, matching metric increase much the same, which indicate that proposed method is robust to rotation and view port change.

5. CONCLUSION

HSI-SFD can not only effectively improve the accuracy and reliability of feature point detection, but also have a good suppression effect on noise, which can effectively improve matching accuracy of feature points in hyperspectral images. At the same time, compared with the existing hyperspectral feature point extraction methods, such as 3D-SIFT, SS-SIFT, etc., the method proposed in this paper can perform multi-band parallel computing and effectively solve the problems of 3D-SIFT-like methods in memory and computation speed.

References

Ait-Aider, O., Berry, F., 2019. A flexible calibration method for the intrinsic and mechanical parameters of panoramic line-scan cameras. *Computer Vision and Image Understanding*, 180, 47–58. https://doi.org/10.1016/j.cviu.2019.01.004.

Al-khafaji, S. L., Zhou, J., Zia, A., Liew, A. W.-C. C., 2018. Spectral-Spatial Scale Invariant Feature Transform for Hyperspectral Images. *IEEE TRANSACTIONS ON IMAGE PRO-CESSING*, 27(2), 837–850.

Bay, H., Ess, A., Tuytelaars, T., Gool], L. V., 2008. Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding*, 110(3), 346 - 359. Similarity Matching in Computer Vision and Multimedia. Christiansen, P. H., Kragh, M. F., Brodskiy, Y., Karstoft, H., 2019. UnsuperPoint: End-to-end Unsupervised Interest Point Detector and Descriptor.

Csurka, G., Dance, C. R., Humenberger, M., 2018. From handcrafted to deep local features. 1–41. http://arxiv.org/abs/1807.10254.

DeTone, D., Malisiewicz, T., Rabinovich, A., 2018. SuperPoint: Self-supervised interest point detection and description. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE.

Dusmanu, M., Rocco, I., Pajdla, T., Pollefeys, M., Sivic, J., Torii, A., Sattler, T., 2019. D2-Net: A Trainable CNN for Joint Detection and Description of Local Features. *Proceedings of the* 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition.

Feng, S., Itoh, Y., Parente, M., Duarte, M. F., 2017. Hyperspectral Band Selection From Statistical Wavelet Models. *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENS-ING*, 55(4), 2111–2123.

Heinly, J., Dunn, E., Frahm, J.-M., 2012. Comparative evaluation of binary features. A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, C. Schmid (eds), *Computer Vision – ECCV 2012*, Springer Berlin Heidelberg, Berlin, Heidelberg, 759–773.

Jin, Y., Mishkin, D., Mishchuk, A., Matas, J., Fua, P., Yi, K. M., Trulls, E., 2020. Image Matching across Wide Baselines: From Paper to Practice. *arXiv*.

Kang, X., Duan, P., Li, S., Benediktsson, J. A., 2018. Decolorization-Based Hyperspectral Image Visualization. *IEEE Transactions on Geoscience and Remote Sensing*, 56(8), 4346–4360. https://doi.org/10.1109/tgrs.2018.2815588.

Lenc, K., Vedaldi, A., 2016. Learning Covariant Feature Detectors. G. Hua, H. Jégou (eds), *Computer Vision – ECCV 2016 Workshops*, Springer International Publishing, Cham, 100–117.

Li, Y., Li, Q., Liu, Y., Xie, W., 2018. A spatial-spectral SIFT for hyperspectral image matching and classification. *Pattern Recognition Letters*, 0, 1–9. https://doi.org/10.1016/j.patrec.2018.08.032.

Lowe, D. G., 2004. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2), 91–110. https://doi.org/10.1023/b:visi. 0000029664.99615.94.

Mukherjee, A., Velez-Reyes, M., Roysam, B., 2009. Interest Points for Hyperspectral Image Data. *IEEE Transactions on Geoscience and Remote Sensing*, 47(3), 748–760.

Ordóñez, Á., Argüello, F., Heras, D. B., Demir, B., 2020. GPUaccelerated registration of hyperspectral images using KAZE features. *Journal of Supercomputing*, 1–15.

Ordóñez, Á., Argüello, F., Heras, D. B., Ordóñez, Á., Argüello, F., Heras, D. B., 2018. Alignment of Hyperspectral Images Using KAZE Features. *Remote Sensing*, 10(5), 756.

Rister, B., Horowitz, M. A., Rubin, D. L., 2017. Volumetric Image Registration From Invariant Keypoints. *IEEE Transactions on Image Processing*, 26(10), 4900–4910. Vedaldi, A., Fulkerson, B., 2008. VLFeat: An open and portable library of computer vision algorithms. http://www.vlfeat.org/.

Zhang, X., Yu, F. X., Karaman, S., Chang, S.-F., 2017. Learning discriminative and transformation covariant local feature detectors. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE.