

## A SIMPLE AND EFFICIENT CROSS-SENSOR RETRIEVAL METHOD FOR RETRIEVING STEREO IMAGES BY MULTISPECTRAL IMAGE

Feifei Peng<sup>1</sup>, Wei Lu<sup>1,\*</sup>, Lili Xu<sup>1</sup>, Xiaoyu Cai<sup>2</sup>

<sup>1</sup> College of Urban and Environmental Sciences, Central China Normal University, Wuhan, China - (feifpeng, luweicnun, xulls)@mail.cnnu.edu.cn

<sup>2</sup> China Mobile Group Design Institute Co., Ltd. Hubei Branch, Wuhan, China - CXY1126710@163.com

### Commission III

**KEY WORDS:** Image Retrieval, Stereo Image, Multispectral Image, Raw Viewing Image, Probability Histogram

#### ABSTRACT:

Some users need to utilize a query multispectral image to quickly locate desired panchromatic stereo images from massive remotely sensed images. A stereo pair or triplet is different with a multispectral image in terms of the viewing number, viewing angle, band, radiometric resolution, spatial resolution, and ability to obtain height information. To perform the cross-sensor retrieval, the orthoimage or digital surface model (DSM) is usually produced from stereo images in a long time, drastically reducing the retrieval efficiency. To achieve a high efficiency, our study explores the potential of the raw viewing images of stereo images to be immediately used in the retrieval. We proposed a simple and efficient cross-sensor retrieval method by doing similarity matching between the query multispectral image and the raw viewing images of stereo images, using probability histograms separately produced from them. Experimental results show that our method outperformed two methods based on the orthoimage in terms of both the retrieval efficiency and precision. Our method handily deals with differences between stereo images and multispectral images, and efficiently achieves the high-accuracy cross-sensor retrieval with no need to produce the orthoimage or DSM.

#### 1. INTRODUCTION

Massive high-resolution stereo images have been acquired by some satellites, such as WorldView-1~4, Cartosat-1/2, ALOS-PRISM, and ZY-3 (Poli and Toutin 2012, Peng, Gong et al. 2017). This has created a surging demand for effective stereo image retrieval methods to locate desired stereo images from massive remotely sensed images (Peng, Wang et al. 2015, Yan, Shaker et al. 2015, Qi, Liu et al. 2017, Peng, Wang et al. 2019). Existing stereo image retrieval methods (Xu, Geng et al. 2014, Chaker, Kaaniche et al. 2015, Peng, Wang et al. 2015, Peng, Luo et al. 2017) require users to submit a query with a stereo pair or triplet, and meet requirements of some users. However, some users only have a multispectral image, which is more easily acquired than stereo images, to submit a query. These methods and existing multispectral image retrieval methods (Quartulli and G. Olaizola 2013, Özkan, Ateş et al. 2014, Demir and Bruzzone 2016, Li, Zhang et al. 2016, Wang, Zhang et al. 2016) have limited applicability to achieve the cross-sensor retrieval. This is because that these methods do not consider differences between stereo images and multispectral images in terms of the viewing number, viewing angle, band, radiometric resolution, and spatial resolution. Stereo images consist of two or three panchromatic images acquired at different off-nadir viewing angles, while a multispectral image usually having blue, green, red and near-infrared bands is acquired at a nadir or near-nadir viewing angle. For these methods, the similarity between a multispectral image and a stereo pair or triplet is done by measuring the similarity between the multispectral image and the orthoimage produced from the stereo pair or triplet, using planar spectral, shape, and texture features (Peng, Wang et al. 2015). The orthoimage production takes much time (Poli and Toutin 2012), and thus would drastically reduce the retrieval

efficiency. Moreover, the re-sampling in the orthoimage production (Aguilar, Agüera et al. 2008, Aguilar, Saldaña et al. 2013) may affect feature extraction, and image retrieval. It is still unknown if the raw viewing panchromatic images of stereo images, rather than the produced orthoimage, could be immediately used in the retrieval. The potential of the raw viewing panchromatic images for the cross-sensor retrieval is explored in this study.

Height information is also usually used by existing stereo image retrieval methods (Peng, Wang et al. 2015, Peng, Luo et al. 2017). The Digital Surface Model (DSM) can be immediately generated from a stereo pair or triplet (Reinartz, d'Angelo et al. 2010, Sirmacek, Taubenboeck et al. 2012, Aguilar, Saldana et al. 2014), but is difficult to be obtained from a multispectral image (Ghamisi, Yokoya et al. 2017, Ghamisi and Yokoya 2018). Recently, some novel approaches simulate the DSM from a single multispectral image with conditional generative adversarial nets (Ghamisi and Yokoya 2018, Mou and Zhu 2018, Amirkolaei and Arefi 2019). These approaches need to establish an image-to-DSM translation rule with several scenes where both DSM and optical data are available, consuming quite a long time. It would be uneconomical to apply height information obtained with these approaches to perform the cross-sensor retrieval. It would take much time and effort to collect the required scenes and then train this approach, drastically reducing the retrieval efficiency. Therefore, existing stereo image retrieval methods using height information have limited applicability to achieve the cross-sensor image retrieval for using a multispectral image to retrieve stereo images.

Existing stereo image retrieval methods using height information can be divided into two categories according to the

\* Corresponding author

height information application way. The first category of methods applies height information alone. Bedoya (2011) measured the similarity between digital elevation models based on landforms. We achieved stereo image retrieval with the Normalized Regional Fractal Co-occurrence Matrix (NRFCM) generated from the stereo-extracted DSM (Peng, Wang et al. 2015). This method achieved low retrieval precision, since different land cover types may be very similar in the DSM (Aguilar, Saldana et al. 2014, Peng, Wang et al. 2015, Ghamisi and Yokoya 2018), e.g., farmland and playgrounds. Xu, Geng et al. (2014) and Chaker, Kaaniche et al. (2015) performed stereo image retrieval with the stereo-extracted disparity information, targeting stereo images acquired by close-range imaging sensors. The disparity information relates to the image spatial resolution and the convergence angle of a stereo pair (Toutin 2004), and cannot be obtained from a multispectral image. Therefore, these methods do not apply to use a query multispectral image to retrieve spaceborne or airborne stereo images.

The second category of stereo image retrieval methods applies both planar and height information. Feng, Ren et al. (2011) employed stereo-extracted disparity to refine results from conventional image retrieval methods in a re-ranking scheme, achieving promising results when retrieving close-range stereo pairs. Chaker, Kaaniche et al. (2015) proposed two methods through univariate and bivariate models in the wavelet-transform domain, combining the visual contents and stereo-extracted disparity information. Xu, Geng et al. (2014) proposed a framework for object-based stereo image retrieval, using both color and stereo-extracted depth to detect objects from images. A generic framework for stereo image retrieval was achieved with a combination of height and planar features (Peng, Wang et al. 2015). A high-accuracy stereo image retrieval method was achieved with height and planar visual word pairs (Peng, Luo et al. 2017), which were separately produced from the stereo-extracted DSM and orthoimage.

## 2. METHODOLOGY

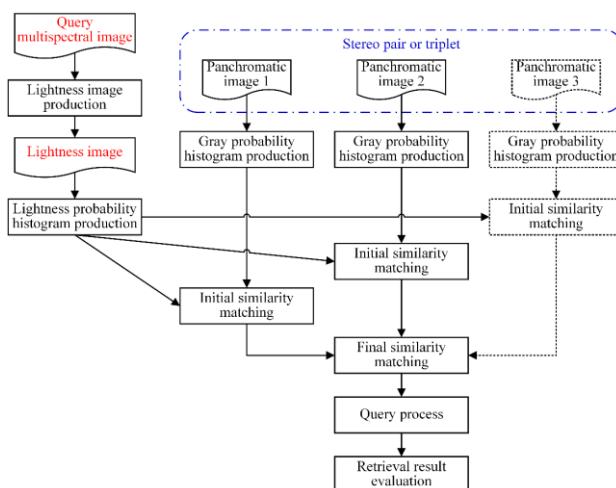


Figure 1. Flowchart of the proposed cross-sensor retrieval method for using a multispectral image to retrieve stereo images.

Figure 1 shows the flowchart of the proposed cross-sensor retrieval method for using a multispectral image to retrieve stereo images. The query multispectral image and stereo images are first separately processed to alleviate differences between them, especially on the band number. For the query multispectral image, a lightness probability histogram is

produced from the lightness image created from the multispectral image. For stereo pairs or triplets, a gray probability histogram is produced for each viewing panchromatic image, creating two or three gray probability histograms. Then, the lightness probability histogram and the gray probability histogram are used to achieve initial similarity matching between the query multispectral image and a viewing image. Final similarity matching between the query multispectral image and a stereo pair or triplet is done based on initial similarity matching results for different viewing images. A query process is done based on final similarity matching, and is followed by a retrieval result evaluation.

### 2.1 Lightness Probability Histogram Production

A lightness image is first produced from the query multispectral image having red, green and blue bands. For each pixel in the multispectral image, its lightness value (Luo, Wang et al. 2015) is obtained as follows:

$$l_{\text{raw}} = (R + G + B) / \sqrt{3}, \quad (1)$$

where,  $R$ ,  $G$ , and  $B$  denote pixel values in red, green, and blue bands, respectively. Then, the lightness value for each pixel is normalized as follows:

$$l = l_{\text{raw}} \times h / (\sqrt{3} \times K_{\text{multi}}), \quad (2)$$

where,  $K_{\text{multi}}$  is determined by the radiometric resolution of the multispectral image.  $h$  determines the range of normalized lightness values, which equals to the range of normalized gray values in panchromatic images. The lightness values of all pixels are calculated in turn, creating a lightness image for the multispectral image.

Then, a lightness probability histogram is created, based on the produced lightness image, for the query multispectral image. The probability histogram (Gonzalez and Woods 2007) has  $h$  bins in total. The probability for each bin is calculated by

$$m(i) = n(i) / \sum_{j=1}^h n(j), \quad 1 \leq i \leq h, \quad 0 \leq m(i) \leq 1, \quad (3)$$

where,  $n(j)$  is the total number of pixels whose normalized lightness values equal to  $j$ .

### 2.2 Gray Probability Histogram Production

For a viewing panchromatic image of stereo images, raw gray values of its pixels are first normalized before the gray histogram production. The normalized gray value for each pixel is calculated as follows:

$$g = g_{\text{raw}} \times h / K_{\text{pan}}, \quad (4)$$

where,  $g_{\text{raw}}$  is the raw gray value of the pixel.  $h$  determines the range of normalized gray values, such as ranging from 1 to  $2^8$  when  $h$  is set to  $2^8$ .  $K_{\text{pan}}$  is determined by the radiometric resolution of the panchromatic image, e.g.,  $2^{11}$  for Worldview-1/2 panchromatic images.

Then, a gray probability histogram is created for the panchromatic image. The probability histogram also has  $h$  bins in total, while the probability for each bin is calculated according to (3). For a stereo pair or triplet, two or three gray probability histograms are separately produced for each viewing image in the same way.

### 2.3 Initial Similarity Matching

Initial similarity matching between the query multispectral image and a viewing image of stereo images is done based on the produced lightness probability histogram and the gray probability histogram. Initial similarity matching is done by

$$s = \sum_{j=1}^h \min(m(j), g(j)), \quad (5)$$

where,  $m(j)$  and  $g(j)$  are the  $j$ th bin of the lightness probability histogram and the gray probability histogram, respectively. For each viewing image of stereo images, initial similarity matching between it and the query multispectral image is done according to (5), obtaining an initial similarity.

### 2.4 Final Similarity Matching

Final similarity matching between the query multispectral image and a stereo pair or triplet is done based on the initial similarity for each viewing image. Final similarity matching is done as follows:

$$S_{final} = \sum_{a=1}^t s_a \times \cos O_a / \sum_{a=1}^t \cos O_a, \quad (6)$$

where,  $s_a$  is the initial similarity between the  $a$ th viewing image and the query multispectral image.  $O_a$  is the off-nadir angle of the  $a$ th viewing image. The parameter  $t$  is set to 2 for stereo pairs, but 3 for stereo triplets.

### 2.5 Cross-sensor Image retrieval

Cross-sensor image retrieval is achieved based on similarity matching between the query multispectral image and stereo images. Users first submit a query multispectral image to start a query process. For each stereo pair or triplet in the dataset, the similarity between it and the query multispectral image is respectively calculated according to (6). Similarities for all stereo pairs or triplets in the dataset are then sorted in descending order, finally returning a similarity ranking list to users.

### 2.6 Retrieval result evaluation

The cross-sensor image retrieval method is evaluated in terms of the retrieval accuracy and efficiency. The retrieval efficiency refers to the total consuming time for the query process. A short time indicates that the retrieval method has a high efficiency. Retrieval precision is a popular index to evaluate retrieval accuracy (Peng, Luo et al. 2017) and, thus, is used in our study. Retrieval precision refers to the percentage of true relevant stereo pairs or triplets, which look like the query multispectral image, in the first several results (e.g., 20) obtained by users.

## 3. RESULTS

### 3.1 Study data

As shown in Figure 2, the study area, having a size of 67.5 square kilometers, is located in Washington, DC, USA. There were a Worldview-1 panchromatic stereo pair, and a Worldview-2 multispectral image for the study area. As can be seen from Table 1, the stereo pair is different with the multispectral image in terms of the band, viewing number, spatial resolution, and observation angles. The complete stereo pair and the complete multispectral image were separately

divided into 3000 small-sized images in the same way. Each small-sized scene has a stereo pair, 300 × 300 pixels in size (for image 1), and a multispectral image, 80 × 80 pixels in size. These small-sized stereo pairs constituted the test dataset, while these multispectral images were for users to choose a query image.

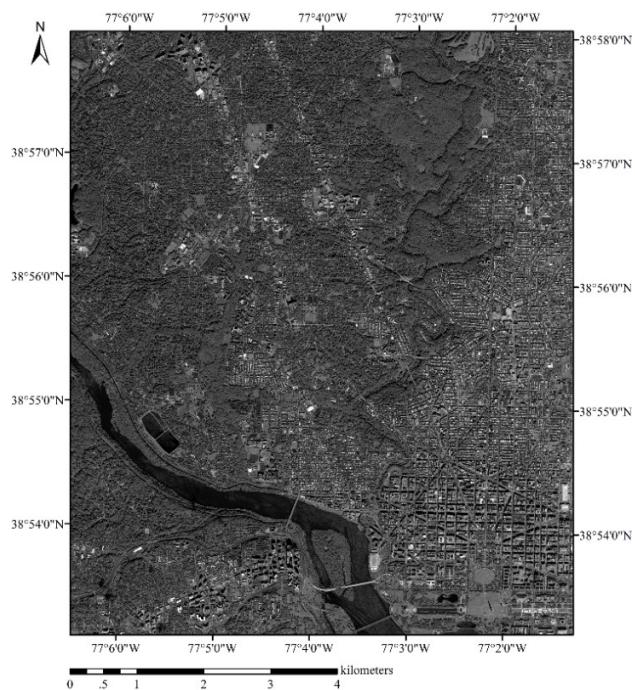


Figure 2. Panchromatic image 1 of the Worldview-1 stereo pair covering the study area.

Table 1. Characteristics of the used stereo pair and multispectral image.

Characteristics	Stereo pair		Multispectral image
	Image 1	Image 2	
Sensor name	Worldview-1		Worldview-2
Band	Panchromatic		Red, Green, Blue
Acquisition date	2012-09-21		2016-08-15
Acquisition time	16:09:14	16:10:17	15:57:30
Radiometric resolution	11 bits	11 bits	11 bits
Ground sample distance (GSD)	0.52 m	0.67 m	1.94 m
Sun azimuth angle	159.7°	160.1°	141.3°
Sun elevation angle	49.8°	49.9°	60.0°
Satellite azimuth angle	1.1°	196.8°	138.2°
Satellite elevation angle	75.0°	54.2°	75.6°
In-track view angle	14.0°	-32.4°	-7.7°
Cross-track view angle	-2.1°	-4.6°	10.1°
Off-nadir view angle	14.2°	32.7°	12.6°

### 3.2 Experiential setting

Our method was compared with four methods in terms of the retrieval accuracy and efficiency. To demonstrate the advantage of using both two raw viewing images of stereo pairs, our method was compared with two methods separately using a single raw viewing image: the image1-based method, and the image2-based method. Our method was also compared with two methods which used the orthoimage generated from stereo images: the ortho-based method, and the NRFCM-based method (Peng, Wang et al. 2015). In order to avoid the influence of the used features or parameters on the retrieval, the image1-based method, the image2-based method, and the ortho-based method also used probability histograms, and were set the same value of  $2^8$  to the parameter  $h$  as our method. The prototype system was implemented using Visual Studio 2010. The testing platform had two Intel(R) Core (TM) i5-4590 3.30 GHz CPUs with 12 GB of memory.

### 3.3 Probability histogram comparison

Figure 3 shows produced probability histograms for the multispectral image and the panchromatic stereo pair covering the study area. The distribution curves of three probability histograms varied with each other in many ways. Specifically, the wave crest of the curve was around 40 for the lightness probability histogram of the multispectral image and the gray probability histogram of the panchromatic image 1, but was around 16 for the gray probability histogram of the panchromatic image 2. There were hardly any values less than 16, and a lot of values more than 96 for the lightness probability histogram of the multispectral image. In contrast, there were hardly any values more than 96, and a number of values less than 16 for the gray probability histograms of the stereo pair, especially for the panchromatic image 2. Overall, three probability histograms were comparable. In comparison of the panchromatic image 2, the gray probability histogram of the panchromatic image 1 was more comparable with the lightness probability histogram of the multispectral image.

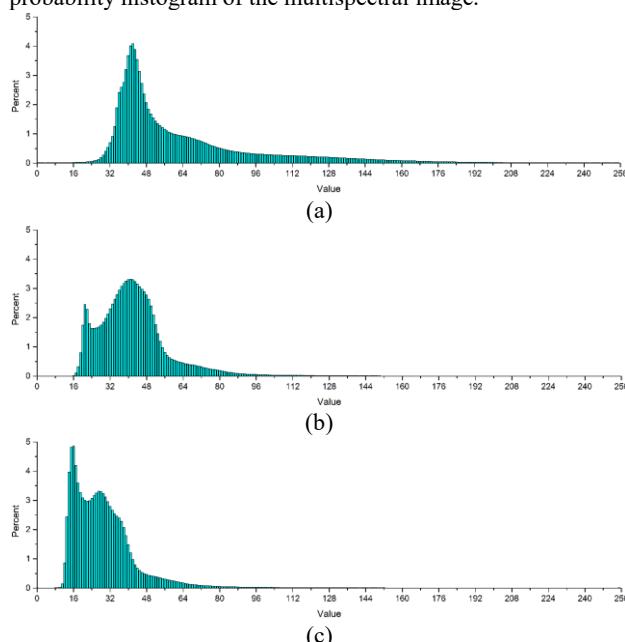


Figure 3. Probability histograms for the multispectral image and the panchromatic stereo pair. (a) Lightness probability histogram for the multispectral image. (b) Gray probability histogram for the panchromatic image 1. (c) Gray probability histogram for the panchromatic image 2.

histogram for the panchromatic image 1. (c) Gray probability histogram for the panchromatic image 2.

### 3.4 Similarity Comparison for Different Viewing Images

Figure 4 shows the scatter plot for two similarities between the query multispectral image and two different viewing panchromatic images of a stereo pair. For a stereo pair, these two similarities were separately represented as horizontal and vertical coordinates of a blue point. In this way, 3000 points were obtained for the test dataset with a query process. As can be seen from Figure 4, two similarities of most stereo pairs are slightly different, while two similarities of few stereo pairs are quite different. This could be due to distinct building shadow and occlusion in two different viewing images of a stereo pair. This results in that the image1-based method and the image2-based method, which separately used the two similarities, obtained different retrieval results and accuracy. Our method used a combination of the two similarities, and achieved higher accuracy.

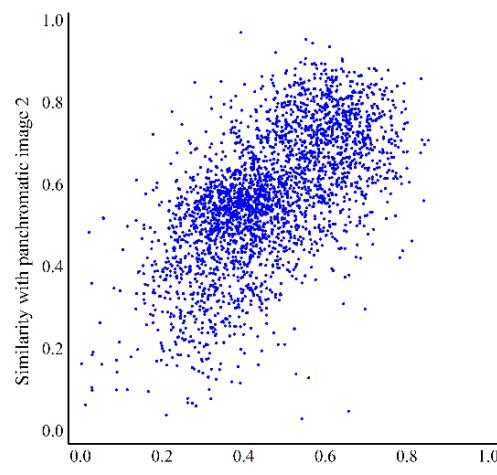


Figure 4. Scatter plot for two similarities between the query multispectral image and two different viewing panchromatic images of a stereo pair.

### 3.5 Cross-sensor image retrieval results

Figure 5 shows the retrieval results of our method and other four methods when using the same query multispectral image. The query multispectral image and the produced lightness image are shown in Figure 5(a). In Figure 5(b)-(f), the seven columns show the seven most similar stereo pairs with descending similarities to the query multispectral, arranged from left to right. For all five methods, there were some stereo pairs which did not look so similar to the query multispectral image, e.g., the third, and seventh stereo pairs for the image1-based method. For the image2-based method and our method, buildings in the first eight stereo pairs were much taller and larger in size than buildings in the query multispectral image. On the whole, the first eight stereo pairs looked similar to the query multispectral image for our method, which is followed by the ortho-based method. Our method achieved better results than two methods separately using a single raw viewing image: the image1-based method, and the image2-based method. The ortho-based method achieved better results than the NRFCM-based method, which also used the orthoimage generated from the raw stereo pair.



Figure 5. Results by our method and other four methods. (a) Query multispectral image and produced lightness image. (b) Image1-based method. (c) Image2-based method. (d) Our method. (e) Ortho-based method. (f) NRFCM-based method.

### 3.6 Retrieval Accuracy Analysis

Figure 6 shows the retrieval accuracy of our method and the other four methods. The precision was calculated according to the first 30 stereo pairs in the similarity ranking list. The image1-based method achieved slightly higher precision than the image2-based method, since that the used panchromatic image 1 was acquired at a small off-nadir viewing angle. The ortho-based method achieved slightly lower precision than the image1-based method and the image2-based method, since that the resampling in the orthoimage production would affect the gray probability histogram production, and further affect similarity matching. The NRFCM-based method, which also was based on the generated orthoimage, achieved the lower precision than the ortho-based method. It indicates that these methods using probability histograms were effective for the cross-sensor retrieval. Our method achieved the highest retrieval precision among the five methods.

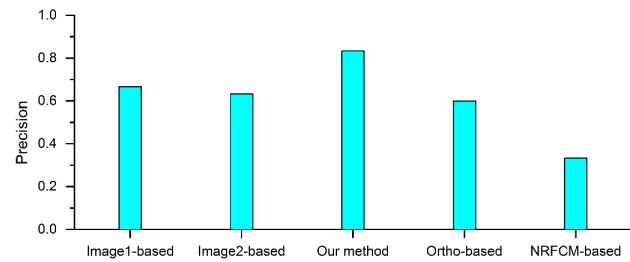


Figure 6. Retrieval accuracy of our method and the other four methods.

### 3.7 Retrieval Efficiency Analysis

Figure 7 shows the retrieval efficiency of our method and the other four methods. Among the five methods, the image1-based method and the image2-based method consumed the least time, since they used a single raw viewing image of a stereo pair. Our method consumed a slightly longer time than the image1-based method and the image2-based method, since our method used both two viewing images of a stereo pair. The ortho-based method and the NRFCM-based method consumed the longest time, since these two methods consumed much time to produce the used orthoimage.

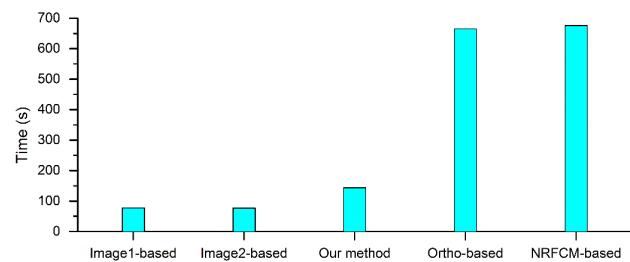


Figure 7. Retrieval efficiency of our method and the other four methods.

## 4. DISCUSSION

Our study demonstrates that the raw viewing panchromatic images of stereo images can be immediately used to achieve the cross-sensor retrieval for using a multispectral image to retrieve stereo images. Our method outperformed two methods based on the generated orthoimage in terms of both the retrieval precision and efficiency. Our method achieved higher precision and a slightly lower efficiency in comparison with two methods using a single raw viewing image of stereo images. For our method, there is no need to like existing stereo image retrieval methods (Xu, Geng et al. 2014, Chaker, Kaaniche et al. 2015, Peng, Wang et al. 2015, Peng, Luo et al. 2017) to consume much time and effort to produce the orthoimage or DSM from stereo images, or to simulate the DSM for the query multispectral image (Ghamisi and Yokoya 2018). Moreover, our method is free from the influence of resampling processing on feature extraction and on image retrieval, since our method uses raw viewing images of stereo images rather than the orthoimage, which requires resampling processing during its production.

To achieve the cross-sensor retrieval, our study handily deals with differences between stereo images and the query multispectral image. Differences on the viewing number and angle are negotiated by doing initial similarity matching between the query multispectral image and each viewing panchromatic image of stereo images, using the viewing angle to determine the weight in final similarity matching. The band

difference is negotiated by producing a lightness image from the multispectral image to make it to be comparable with a viewing panchromatic image. The radiometric resolution difference is negotiated by normalizing raw lightness values and gray values to the same value range. The spatial resolution difference is negotiated by separately producing probability histograms for normalized gray values and lightness values. The ability difference to obtain height information is negotiated by using planar information alone to achieve the retrieval.

Some limitations of our method need to be addressed in future work. Our method uses pixel-based gray values and lightness values to produce probability histograms for doing similarity matching, and thus would suffer from the influence of building shadow and occlusion in images. The influence would be alleviated by using object-based or affine invariant features to do similarity matching. Besides, the influence of the parameter  $h$  on the gray or lightness value normalization, and further on image retrieval will be investigated. Our method will be tested with more stereo images, especially stereo triplets, acquired by more sensors.

## 5. CONCLUSIONS

We proposed a simple and efficient method for using a query multispectral image to retrieve stereo images, utilizing all raw viewing panchromatic images of stereo images. Experimental results show that our method outperformed two methods based on the generated orthoimage in terms of both the retrieval precision and efficiency. Our method achieves a high retrieval precision of 0.8, and a high efficiency. Our method is helpful for users to utilize a multispectral image to quickly and accurately locate desired stereo images from massive spaceborne or airborne stereo pairs and triplets. Our study handily deals with differences between stereo images and multispectral images in terms of the viewing number, viewing angle, band, radiometric resolution, spatial resolution, and ability to obtain height information. Our study demonstrates the possibilities of the raw viewing images of stereo images to be immediately used to achieve the cross-sensor retrieval, with no need to consume much time and effort to produce the orthoimage or DSM. Our method is free from the influence of resampling processing on feature extraction and on image retrieval, and could be extended by using object-based or affine invariant features to achieve higher retrieval precision. In future work, the influence of building shadow and occlusion on the retrieval, the influence of the parameter  $h$ , and the effectiveness of our method for stereo triplets will be investigated.

## ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China under Grant 41701511, the Hubei Provincial Natural Science Foundation of China under Grant 2017CFB278, the Key Laboratory for National Geographic State Monitoring of National Administration of Surveying, Mapping and Geoinformation under Grant 2018NGCMZD03, and the Fundamental Research Funds for the Central Universities under Grant CCNU20TS033.

## REFERENCES

Aguilar, M. A., F. Agüera, F. J. Aguilar and F. Carvajal, 2008. Geometric accuracy assessment of the orthorectification process from very high resolution satellite imagery for Common

Agricultural Policy purposes. *International Journal of Remote Sensing* 29(24), 7181-7197.

Aguilar, M. A., M. d. M. Saldaña and F. J. Aguilar, 2013. Assessing geometric accuracy of the orthorectification process from GeoEye-1 and WorldView-2 panchromatic images. *International Journal of Applied Earth Observation and Geoinformation* 21(0), 427-435.

Aguilar, M. A., M. d. M. Saldana and F. J. Aguilar, 2014. Generation and Quality Assessment of Stereo-Extracted DSM from GeoEye-1 and WorldView-2 Imagery. *IEEE Transactions on Geoscience and Remote Sensing*. 52(2), 1259-1271.

Amirkolaei, H. A. and H. Arefi, 2019. Height estimation from single aerial images using a deep convolutional encoder-decoder network. *ISPRS Journal of Photogrammetry and Remote Sensing* 149, 50-66.

Bedoya, E., 2011. Semantic assessment of similarity between raster elevation datasets. *Rev. Fac. Ing.-Univ. Antioquia* 59(59), 37-46.

Chaker, A., M. Kaaniche and A. Benazza-Benyahia, 2015. Disparity based stereo image retrieval through univariate and bivariate models. *Signal Processing: Image Communication* 31(0), 174-184.

Demir, B. and L. Bruzzone, 2016. Hashing-Based Scalable Remote Sensing Image Search and Retrieval in Large Archives. *IEEE Transactions on Geoscience and Remote Sensing* 54(2), 892-904.

Feng, Y., J. Ren and J. Jiang, 2011. Generic framework for content-based stereo image/video retrieval. *Electronics Letters* 47(2), 97-98.

Ghamisi, P. and N. Yokoya, 2018. IMG2DSM: Height Simulation From Single Imagery Using Conditional Generative Adversarial Net. *IEEE Geoscience and Remote Sensing Letters* 15(5), 794-798.

Ghamisi, P., N. Yokoya, J. Li, W. Liao, S. Liu, J. Plaza, B. Rasti and A. Plaza, 2017. Advances in Hyperspectral Image and Signal Processing: A Comprehensive Overview of the State of the Art. *IEEE Geoscience and Remote Sensing Magazine* 5(4), 37-78.

Gonzalez, R. C. and R. E. Woods, 2007. Digital Image Processing (3rd Edition), Addison Wesley.

Li, Y., Y. Zhang, C. Tao and H. Zhu, 2016. Content-Based High-Resolution Remote Sensing Image Retrieval via Unsupervised Feature Learning and Collaborative Affinity Metric Fusion. *Remote Sensing* 8(9), 709.

Luo, H., L. Wang, Z. Shao and D. Li, 2015. Development of a multi-scale object-based shadow detection method for high spatial resolution image. *Remote Sensing Letters* 6(1), 59-68.

Mou, L. and X. X. Zhu, 2018. IM2HEIGHT: Height estimation from single monocular imagery via fully residual convolutional-deconvolutional network. *arXiv preprint arXiv:1802.10249*.

Özkan, S., T. Ateş, E. Tola, M. Soysal and E. Esen, 2014. Performance Analysis of State-of-the-Art Representation Methods for Geographical Image Retrieval and Categorization. *IEEE Geoscience and Remote Sensing Letters* 11(11), 1996-2000.

Peng, F., J. Gong, L. Wang, H. Wu and P. Liu, 2017. A New Stereo Pair Disparity Index (SPDI) for Detecting Built-Up

Areas from High-Resolution Stereo Imagery. *Remote Sensing* 9(6), 633.

Peng, F., J. Luo, W. Gaoqiang and K. Qi, 2017. Stereo Image Retrieval Using Height and Planar Visual Word Pairs. *IEEE Geoscience and Remote Sensing Letters* 14(11), 2082-2086.

Peng, F., L. Wang, J. Gong and H. Wu, 2015. Development of a Framework for Stereo Image Retrieval With Both Height and Planar Features. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 8(2), 800-815.

Peng, F., L. Wang, S. Zou, J. Luo, S. Gong and X. Li, 2019. Content-based search of earth observation data archives using open-access multitemporal land cover and terrain products. *International Journal of Applied Earth Observation and Geoinformation* 81, 13-26.

Poli, D. and T. Toutin, 2012. Review of developments in geometric modelling for high resolution satellite pushbroom sensors. *The Photogrammetric Record* 27(137), 58-73.

Qi, K., W. Liu, C. Yang, Q. Guan and H. Wu, 2017. Multi-Task Joint Sparse and Low-Rank Representation for the Scene Classification of High-Resolution Remote Sensing Image. *Remote Sensing* 9(1), 10.

Quartulli, M. and I. G. Olaizola, 2013. A review of EO image information mining. *ISPRS Journal of Photogrammetry and Remote Sensing* 75(0), 11-28.

Reinartz, P., P. d'Angelo, T. Krauß, D. Poli, K. Jacobsen and G. Buyukalih, 2010. Benchmarking and quality analysis of dem generated from high and very high resolution optical stereo satellite data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XXXVIII*.

Sirmacek, B., H. Taubenboeck, P. Reinartz and M. Ehlers, 2012. Performance Evaluation for 3-D City Model Generation of Six Different DSMs From Air- and Spaceborne Sensors. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 5(1), 59-70.

Toutin, T., 2004. Review article: Geometric processing of remote sensing images: models, algorithms and methods. *International Journal of Remote Sensing* 25(10), 1893-1924.

Wang, Y., L. Zhang, X. Tong, L. Zhang, Z. Zhang, H. Liu, X. Xing and P. T. Mathiopoulos, 2016. A Three-Layered Graph-Based Learning Approach for Remote Sensing Image Retrieval. *IEEE Transactions on Geoscience and Remote Sensing* 54(10), 6020-6034.

Xu, X., W. Geng, R. Ju, Y. Yang, T. Ren and G. Wu, 2014. OBSIR: Object-based stereo image retrieval. *2014 IEEE International Conference on Multimedia and Expo (ICME)*, China.

Yan, W. Y., A. Shaker and N. El-Ashmawy, 2015. Urban land cover classification using airborne LiDAR data: A review. *Remote Sensing of Environment* 158(0), 295-310.