The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLII-2/W7, 2017
ISPRS Geospatial Week 2017, 18–22 September 2017, Wuhan, China

A Robust Descriptor Based on Spatial and Frequency Structural Information for Visible and Thermal Infrared Image Matching

Zhitao Fu [a,b], Qianqing Qin [a], Chun Wu [a], Yunpeng Chang [a], Bin Luo [a,*]

[a] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China - (zhitaofu,cugwhu,c1571720262)@126.com, qqqin@lmars.whu.edu.cn, luob@whu.edu.cn
[b] Information and Network Centre, Yunnan Minzu University, Kunming, China

Commission III, WG III/6

KEY WORDS: Multimodal Descriptor, Visible and Thermal Infrared Image Matching, Spatial Structural Information, Frequency Structural Information, Multi-spatial Regions Division.

ABSTRACT:

Due to the differences of imaging principles, image matching between visible and thermal infrared images still exist new challenges and difficulties. Inspired by the complementary spatial and frequency information of geometric structural features, a robust descriptor is proposed for visible and thermal infrared images matching. We first divide two different spatial regions to the region around point of interest, using the histogram of oriented magnitudes, which corresponds to the 2-D structural shape information to describe the larger region and the edge oriented histogram to describe the spatial distribution for the smaller region. Then the two vectors are normalized and combined to a higher feature vector. Finally, our proposed descriptor is obtained by applying principal component analysis (PCA) to reduce the dimension of the combined high feature vector to make our descriptor more robust. Experimental results showed that our proposed method was provided with significant improvements in correct matching numbers and obvious advantages by complementing information within spatial and frequency structural information.

1. INTRODUCTION

Image matching is a very important process in image registration, image fusion, image mosaic and changing detection, just to mention a few. The points to be matched are described through their surrounding information providing rich representations. The goal of this process is to design good descriptor. The descriptor should have good distinguishability from other similar regions of the feature points. The descriptor should also be robust to different image transformations. In this paper, we focus on designing a robust descriptor which is robust against non-linear intensity variation between visible image and thermal infrared images, both of which belong to different modalities and different spectra.

Because of their robustness to geometric and illumination differences, most of the matching algorithms are based on local invariant features. Feature detection, feature description and feature matching are three main steps in these algorithms. The most familiar and popular matching method is SIFT (Scale Invariant Feature Transform) (Lowe, 2004), which uses the distribution of gradients for description. Although the SIFT algorithm can give good matching results for visible image pairs, the percentage of matches becomes worse for multi-spectral or multi-modal image pairs for their non-linear intensity differences, particularly in visible image and thermal infrared image pairs.

To cope with such non-linear pixel intensity differences, several modified methods to SIFT have been proposed. A gradient orientation modification SIFT (GOM-SIFT) descriptor was proposed to limit the gradient orientation within $(0, \pi]$ (Yi, 2008). Orientation restricted SIFT (OR-SIFT) was proposed to combine the SIFT descriptor elements in the opposite orientation directions (Vural, 2009). A normalized gradients SIFT (NG-SIFT) descriptor was presented to compute descriptors using a normalized gradient (Saleem, 2013). These modified SIFT methods could work for image registration that the grey level intensities of the pixels in the same region between multispectral images are quite different, inverse better (Saleem, 2014).

Shape context (Belongie, 2000) and Edge Oriented Histogram Descriptor (EHD) (Aguilera, 2012 and Mouats, 2013) are very similar to SIFT descriptor, which computes a histogram based on the gradient distribution around the detected regions. However，both of them computed the histogram by describing the edge distribution around the detected region. Shape context was defined as a joint histogram of the angle and log distance of a point relative to other points. EHD only utilized the edge points which was combined with the spatial distribution of four directional edges and one non-directional edge (5 bins totally). The EHD descriptor was performed better than SIFT, GOM-SIFT and OR-SIFT for multispectral images matching (Aguilera, 2012). A Local Self-Similarity (LSS) descriptor was proposed to capture internal geometric layouts through accounting for small local affine deformations (Shechtman el at., 2007). The self-similarity of colour, edges, repetitive patterns and complex textures were captured in a single unified way. The modified LSS descriptors were applied to multispectral image matching (Kim, 2014 and Sedaghat, 2015) and image registration (Ye, 2014).

* Corresponding author

However, these descriptors are not universally applicable. The combination of different descriptors is hence a natural choice. Some researchers paid more attention to the composition of different descriptors. In order to satisfy the multimodal image registration, a feature-matching strategy, which described the correspondence of the features by combining feature similarity and spatial consistency was proposed (Wen, 2008). Point structured information and a composited descriptor was utilized to combine Daisy descriptor with an improved shape context descriptor to complete multisource image registration (Shi, 2014). A frequency based corner and edge detector combining to EHD between visible and thermal infrared images was presented, which was based on the combination of the spatial shape information and frequency domain named phase congruency around the detected points to obtain more correspondences in multimodal scenario (Mouats, 2013 and 2015). However, the number of correct points was still low. The Log Gabor histogram descriptor (LGHD) used multi-scale and multi-oriented magnitudes of Log-Gabor filters to describe the neighbourhood feature points by combining frequency and spatial information, and the spatial information corresponded to the region around the detected point, which was divided into 4*4 sub regions for more discrimination (Aguilera, 2015). LGHD outperformed than other typical descriptors in correct matching numbers. However, due to discarding the phase information, the accuracy rate of the correct number was low. It would be a good choice to combine EHD and LGHD to improve the accuracy and repeatability of the matching numbers.

In this paper, multi-spatial regions division concept is introduced firstly, and then the LGHD descriptor and the EHD descriptor and established and combined to a higher feature vector. The LGHD is described by using the histogram of oriented magnitudes, which corresponds to the 2-D structural shape information to describe the larger region and the edge oriented histogram to describe the spatial distribution of the smaller region. Considering the expensive matching cost, we apply Principle Component Analysis (PCA) to the feature vector to obtain our proposed descriptor.

The main contribution of this paper includes two aspects. One is to combine spatial and frequency structural in-formation using multi-spatial regions division concept, the other is to apply PCA to make our proposed descriptor more compact and robust. The rest of this paper is organized as follows. The methodology is described in Section 2. Evaluation Setup is proposed in Section 3, followed by experimental results in Section 4. Conclusions are presented in Section 5.

2. PROPOSED APPROACH

The differences of imaging principles between sensors make the correlation between pixels different and non-linear intensity variations between a pair of visible and thermal infrared images. However, in spite of intensity differences, the edges and the global appearance of the shape of the objects in the scenes tends to be less changed. The EOH descriptor (EHD) describes the spatial edge distribution around point of interest in spatial domain, while LGHD draws the distribution of high frequency components corresponding to frequency domain. We can predict that a descriptor based on spatial and frequency structural information would be more robust for matching non-linear visible and thermal infrared images, which is the main idea behind the proposed approach.

The current work is based on EHD descriptor and LGHD descriptor. For each region around point of interest, the size of S*S is divided into 4*4 sub regions. Within each sub region, a 5-bin edge oriented histogram is calculated by acquiring the strongest value for one of five different oriented *Sobel* Filters, all EOHs in each sub region combined together for 80-bin EHD descriptor. In contrast, LGHD uses the 6-oriented and 4-scale Log Gabor filters to compute 6-bin histogram of magnitudes (HOM) in each sub region of each scale. After joining all HOMs in each region and all scales, a 384-bin feature vector is obtained. The region size of both methods is suggested to 80*80 (Aguilera, 2015).

Our proposed approach is based on a combination the EHD descriptor and LGHD descriptor. We present three main steps to obtain our robust descriptor. The flowchart of proposed method is shown in Figure 1.

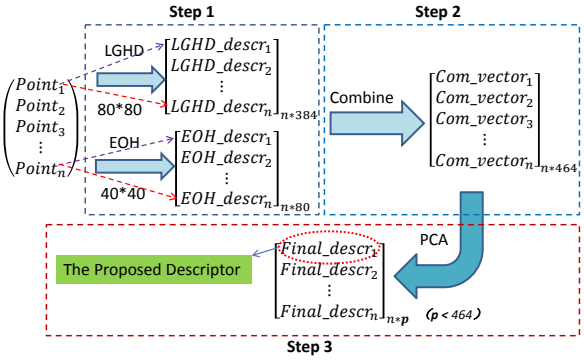


Figure 1. The flowchart of proposed approach

In Step 1, in order to make our descriptor more discriminative, we apply two spatial regions division to the process of description of EHD and LGHD. The region size of EHD is assigned to 40*40 while LGHD is preserved as 80*80.

In Step 2, both EHD and LGHD are normalized to unit length before being combined to a high feature vector. One reason is to cancel contrast change, the other and more important is to satisfy the requirement of combination. The EHD descriptor is 80-bin vector, and LGHD is 384-bin vector, the combined vector sums up to 464 bins.

In Step 3, we apply PCA, which is one of the well-known techniques to reduce dimensionality and make the characteristic expression more compact, to compress the dimensionality of the combined vectors. The compressed vector is regarded as our final descriptor. The criterion for choosing p is to select a cumulative percentage of total variation which one desires that the selected eigenvalues contribute, and 85% is suggested in our experiments.

3. EXPERIMENTAL SETUP

We will introduce the details of experimental settings, including the feature detector, descriptors, datasets and evaluation criteria in this section.

3.1 Feature Detector and Feature Descriptors

The phase congruency corner detector (Kovesi, 2003) is used in our experiments to detect stable points of interest and their surrounding support regions. The phase congruency corner detector is invariant to contrast and illumination. Therefore, it is suitable for visible and thermal infrared image matching

(Mouats, 2013). We use non-maximum suppression method to get the phase congruency corner values, which are sorted in decrease. We select 400 maximum interest points in visible image and the same number of points corresponding to thermal image.

In our experiments, we utilize four feature descriptors, including *SIFT* (Lowe, 2004), *EHD* (Aguilera, 2012), *PCEHD* (Mouats, 2013), and *LGHD* (Aguilera, 2015), to compare with our proposed descriptor. The patches size of EHD, PCEHD and LGHD is the same (80*80) for fair evaluation. The patch size of SIFT descriptor is given by default. Euclidean distance is used to measure the dissimilarity between two regions under each descriptor. The Nearest Neighbour Distance Ratio (*NNDR*) (Lowe 2004) is adopted to complete ambiguity rejection. The NNDR threshold is assigned from 0.45 to 1.0 at intervals of 0.05 in our experiments. A match is correct when it is within 3 pixels of its predicted position.

3.2 Datasets and Evaluation Criteria

The performance of our approach is evaluated on 44 pairs of visible and thermal infrared images acquired from CVC dataset (Aguilera, 2015). These pairs of images include various types of scenes, and different numbers of common objects. They jointly serve as a good test bed for our performance evaluation. One pair of visible and thermal infrared (TIR) images is shown in Figure 2.

Figure 2. One pair of visible (left) and TIR (right) images

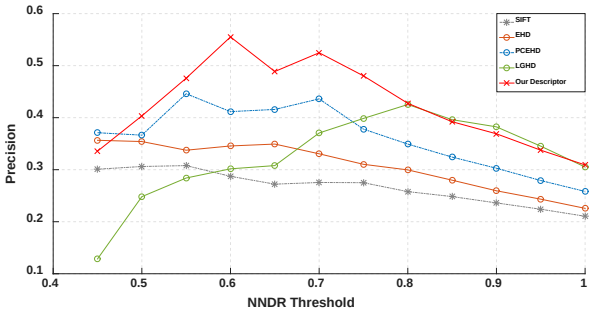
To evaluate the performance of our proposed approach, precision and recall are taken into account (Mikolajczyk, 2005 and Zhao, 2009). Recall is the fraction of correct correspondences that are detected, while precision is the fraction of detected correspondences that are correct. They are defined as

$$Precision = \frac{n_{cm}}{n_{cm} + n_{fm}}; Recall = \frac{n_{cm}}{n_{cp}} \quad (1)$$

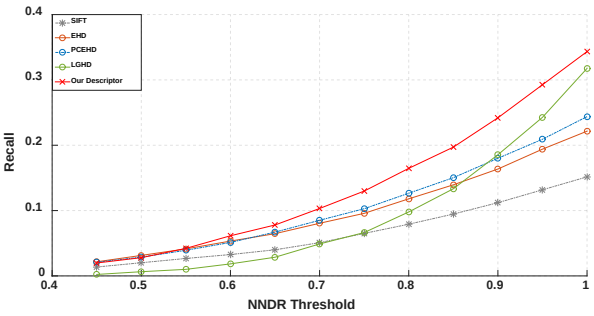
where n_{cm} and n_{fm} are the numbers of correct matching and the numbers of false matching, n_{cp} stands truth number of matching regions between the pair of images. The performance of each descriptor can then be illustrated by a precision-recall curve on visible and thermal infrared image pair on our dataset.

4. EXPERIMENTAL RESULTS

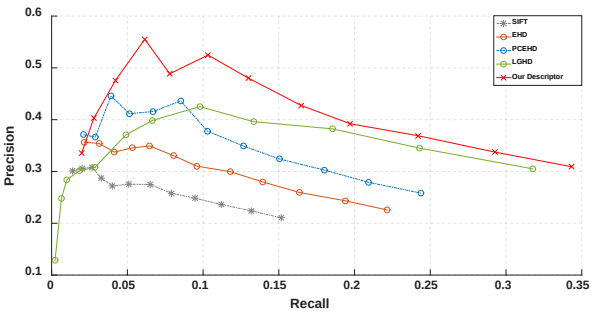
The results of proposed approach compared with other four methods are shown in Figure 3, where Figure 3(a) indicates the average precision results of 44-pair images, Figure 3(b) represents the average recall results of different descriptors, and both NNDR threshold ranges from 0.45 to 1.0 at intervals of 0.05. The average precision-recall curves of different methods are shown in Figure 3(c).



(a) The average precision results of different methods



(b) The average recall results of different methods



(c) The precision-recall curves of different methods

Figure 3. The average results of 44-pair multimodal images in CVC dataset with different descriptors

It is obvious that our proposed descriptor performed best according to Figure 3(a). All edge oriented histogram-based descriptor, including EHD and PCEHD, displayed better than LGHD and SIFT when the NNDR threshold is low. As the NNDR threshold increases to 1.0, the precision of all descriptors decrease. However, our proposed descriptor and LGHD represented better than other descriptors. In general, our proposed descriptor outperformed other four typical descriptors. We can see that the average recall curve of our descriptor is higher than other descriptors in Figure 3(b). As the NNDR increased, the LGHD and other three descriptors illustrated better and better, especially our descriptor and LGHD descriptor. The precision-recall curves of different descriptors are also shown that our proposed descriptor is obviously more excellent than other descriptors in Figure 3(c). It is evident from the results that our proposed descriptor obtained the best performance when compared with all other descriptors evaluated in our work. Due to the combination spatial information and frequency information, and multi-spatial regions division, both the precision and the recall curves of our descriptor are higher than other descriptors, which mean our descriptor is more discriminative and more robust for matching visible and thermal infrared images.

5. CONCLUSIONS

In this paper, a robust and discriminative descriptor for visible and thermal image matching is proposed to address the issue of correlation between pixels different. Three steps are shown to our proposed approach. The multi-spatial regions division concept is introduced to make our descriptor more discriminative, and combining spatial and frequency structural information made our descriptor more robust. After dimensionality reduction to combined vectors, our proposed descriptor is more compact and more robust, and the matching efficiency is improved. We have demonstrated our proposed method with *SIFT*, *EHD*, *PCEHD* and *LGHD* on various scenes of visible and thermal infrared images. The experimental results showed that our proposed descriptor outperformed other methods. It is worth to note that performance of our descriptor may decline if the images have few structure or shape information, because our descriptor depends on the spatial and frequency structural properties. An image enhancement may be useful for image matching. A more thorough evaluation will be addressed using more visible and thermal infrared images in our future work.

ACKNOWLEDGEMENTS

This work was supported part by the National Natural Science Foundation of China under grant No.6157011347 and 61261130587, and the National High Technology Research and Development Program of China (863 Program) under grant No.2013AA122301 and 2012AA12A305. It was also supported by the Youth Foundation of Yunnan Minzu University under grant No.2016QN03 and the Scientific Research Fund of Yunnan Provincial Education Department under grant No.2017ZZX088.

REFERENCES

A. Sedaghat and H. Ebadi, 2015. Distinctive Order Based Self-Similarity descriptor for multi-sensor remote sensing image matching, ISPRS Journal of Photogrammetry and Remote Sensing, vol. 108, pp. 62-71.

C. A. Aguilera, A. D. Sappa, and R. Toledo, 2015. LGHD: A feature descriptor for matching across non-linear intensity variations, in Image Processing (ICIP), 2015 IEEE International Conference, pp. 178-181.

C. Aguilera, F. Barrera, F. Lumbreras, A. D. Sappa, and R. Toledo, 2012. Multispectral Image Feature Points, Sensors, vol. 12, pp. 12661-72.

D. G. Lowe, 2004. Distinctive Image Features from Scale-Invariant Keypoints, International Journal of Computer Vision, vol. 60, pp. 91-110.

E. Shechtman and M. Irani, 2007. Matching Local Self-Similarities across Images and Videos, IEEE Conference in Computer Vision and Pattern Recognition, CVPR '07, pp. 1-8.

G. J. Wen, J. J. Lv, and W. X. Yu, 2008. A High-Performance Feature-Matching Method for Image Registration by Combining Spatial and Similarity Information, IEEE Transactions on Geoscience & Remote Sensing, vol. 46, pp. 1266-1277.

K. Mikolajczyk and C. Schmid, 2005. A performance evaluation of local descriptors, IEEE Transactions on Pattern Analysis & Machine Intelligence, vol. 27, pp. 1615, 2005.

M. F. Vural, Y. Yardimci, and A. Temizel, 2009. Registration of multispectral satellite images with Orientation-Restricted SIFT, in Geoscience and Remote Sensing Symposium, IEEE International, igarss, pp. III-243-III-246.

P. Kovesi, 2003. Phase Congruency Detects Corners and Edges, in DICTA, pp. 309-318.

Q. Shi, G. Ma, F. Zhang, W. Chen, Q. Qin, and D. Huang, 2014. Robust Image Registration Using Structure Features, IEEE Geoscience & Remote Sensing Letters, vol. 11, pp. 2045-2049.

S. Belongie, J. Malik, and J. Puzicha, 2000. Shape Context: A new descriptor for shape matching and object recognition, pp. 831-837.

S. Kim, S. Ryu, B. Ham, J. Kim, and K. Sohn, 2014.Local self-similarity frequency descriptor for multispectral feature matching, in Image Processing (ICIP), 2014 IEEE International Conference, pp. 5746-5750.

S. Saleem and R. Sablatnig, 2013. A Modified SIFT Descriptor for Image Matching under Spectral Variations, Lecture Notes in Computer Science, vol. 8156, pp. 652-661.

S. Saleem and R. Sablatnig, 2014. A robust sift descriptor for multispectral images, IEEE signal processing letters, vol. 21, pp. 400-403.

T. Mouats and N. Aouf, 2013. Multimodal stereo correspondence based on phase congruency and edge histogram descriptor, in International Conference on Information Fusion, pp. 1981-1987.

T. Mouats, N. Aouf, A. D. Sappa, C. Aguilera, and R. Toledo, 2015. Multispectral Stereo Odometry, IEEE Transactions on Intelligent Transportation Systems, vol. 16, pp. 1210-1224.

W. L. Zhao and C. W. Ngo, 2009. Scale-rotation invariant pattern entropy for keypoint-based near-duplicate detection, IEEE Transactions on Image Processing, vol. 18, pp. 412-23.

Y. Ye and J. Shan, 2014. A local descriptor based registration method for multispectral remote sensing images with non-linear intensity differences, ISPRS Journal of Photogrammetry and Remote Sensing, vol. 90, pp. 83-95.

Z. Yi, C. Zhiguo, and X. Yang, 2008. Multi-spectral remote image registration based on SIFT, Electronics Letters, vol. 44, pp. 107-108.