ISPRS Journal of Photogrammetry and Remote Sensing 142 (2018) 205–221

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

A local phase based invariant feature for remote sensing image matching

Yuanxin Ye[a], Jie Shan[b], Siyuan Hao[c,*], Lorenzo Bruzzone[d], Yao Qin[e]

[a] Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu 610031, China
[b] School of Civil Engineering, Purdue University, West Lafayette, IN 47907, USA
[c] College of Information and Control Engineering, Qingdao University of Technology, Qingdao 266520, China
[d] Department of Information Engineering and Computer Science, University of Trento, Trento 38123, Italy
[e] College of Electronic Science and Engineering, National University of Defense Technology, Changcha 410073, China

ARTICLE INFO

Keywords:
Image matching
Remote sensing images
Local invariant features
Radiometric differences

ABSTRACT

Local invariant features from computer vision community have recently been widely applied to the matching of remote sensing images. However, these features are mainly designed to handle geometric distortions, and are sensitive to complex radiometric differences between multisensor images. To address this issue, this paper proposes an effective local invariant feature that is sufficiently robust to both geometric and radiometric changes. The proposed feature is built based on the phase congruency model that is invariant to illumination and contrast variation. It consists of a feature detector named MMPC-Lap and a feature descriptor named local histogram of orientated phase congruency (LHOPC). MMPC-Lap is constructed by using the minimum moment of phase congruency for feature detection with an automatic scale location technique, which is used to detect stable keypoints in image scale space. Subsequently, LHOPC derives the feature descriptor for a keypoint by utilizing an extended phase congruency feature with an advanced descriptor configuration. Finally, correspondences are achieved by evaluating the similarity of the feature descriptors. The proposed MMPC-Lap and LHOPC have been evaluated under different imaging conditions (spectral, temporal, and scale changes). The results obtained on a variety of remote sensing images demonstrate its excellent performance with respect to the state-of-the-art local invariant features, especially for cases where there are complex radiometric differences.

1. Introduction

Image matching aims to detect correspondences or control points (CPs) between two or more images (Ye et al., 2017b). It is a key step for subsequent remote sensing image processing such as image alignment (Cole-Rhodes et al., 2003), 3D reconstruction(Ahmadabadian et al., 2013), image stitching (Li et al., 2015), and change detection (Bruzzone and Bovolo, 2013). Due to their different imaging mechanisms and spectral properties, multisensory or multispectral remote sensing images usually have significant geometric and radiometric distortions (see Fig. 1), which make automatic matching of such images challenging.

Local invariant features have a rapid development in computer vision field, and have got considerable attention from researchers of remote sensing field. Some popular local invariant features, such as the scale-invariant feature transform (SIFT) (Lowe, 2004), the Speed Up Robust Feature (SURF) (Bay et al., 2008), and shape context (Belongie et al., 2002), have been applied to remote sensing image matching. However, these features mainly address geometric distortions such as scale and rotation changes, but are sensitive to complex radiometric differences between images (Kelman et al., 2007; Ye et al., 2017a). Therefore, developing a local invariant feature sufficiently robust to geometric and radiometric changes for remote sensing image matching is highly desirable.

2. Related work

Existing matching methods can be roughly divided into two categories: intensity-based and feature-based. Intensity-based methods detect CPs between images based on some similarity measures, such as normalized cross correlation, mutual information, and matching by tone mapping (Hel-Or et al., 2014). These methods require initial matching positions, and are vulnerable to geometric distortions. Feature-based methods first detect the salient features between images, and then construct descriptors to depict the features' properties and then use the similarity of the feature descriptors to achieve correspondences. Different from intensity-based methods, feature-based method usually does not rely on initial matching conditions, and are more robust to geometric and radiometric differences. Nowadays, local invariant features have been widely used for remote sensing image matching

* Corresponding author.
E-mail address: lemonbananan@163.com (S. Hao).

https://doi.org/10.1016/j.isprsjprs.2018.06.010
Received 9 January 2018; Received in revised form 11 June 2018; Accepted 15 June 2018

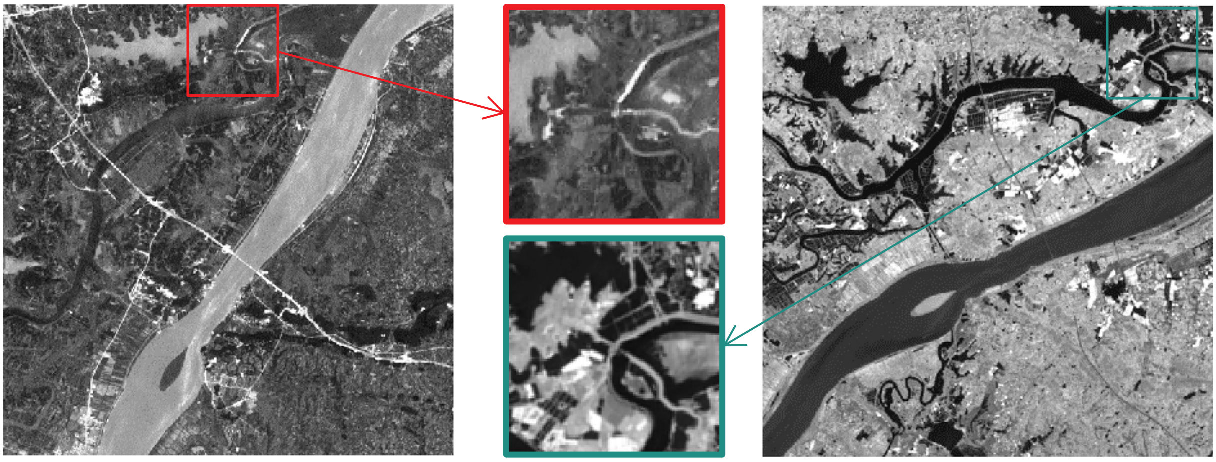

Fig. 1. An example of geometric and radiometric distortions between multispectral remote sensing images. (a) TM band 1 (visible). (b) TM band 4 (infrared). The two images have significant rotation and radiometric differences. The radiometric differences are particularly significant as shown in the enlarged patches, which makes correspondence detection difficult, even by visual inspection.

because of their robustness to geometric and illumination changes. In general, local invariant features are composed of a feature detector and feature descriptor.

Feature detectors are algorithms that detect the distinctive features between images, such as corners, blobs, and image patches. In the past several decades, feature detectors have undergone a rapid development, during which many different detectors were proposed. One of the first detectors was proposed by Moravec, which was able to detect interest points (e.g., corners) by using the sum of the squared differences of intensities within local image regions (Moravec, 1980). This operator offers fast computation, but it has a low repeatability rate and is sensitive to noise. In order to address the limitations of the Moravec operator, the Harris detector was developed based on an auto-correlation function of gradient information (Harris and Stephens, 1988). This detector is more accurate and offers a higher repeatability rate. However, the Harris detector is vulnerable to image scale changes (Schmid et al., 2000) and cannot be applied to match images with large scale differences. To address this issue, Lindeberg studied image scale methodology thoroughly, and found that the Laplacian of Gaussian (LoG) detector could reflect the properties of local characteristic scales (Lindeberg, 1994, 1998). This finding established the foundation for the development of scale invariant feature detection. Subsequently, by combining LoG with the Harris and Hessian detectors, Mikolajczyk et al. proposed two feature detectors with scale invariance, which are named Harris-Laplace (Harris-Lap) and Hessian-Laplace (Hessian-Lap), respectively (Mikolajczyk and Schmid, 2004; Mikolajczyk et al., 2005). Lowe developed the differences of Gaussian (DoG) detector to extract interest points with scale invariance (Lowe, 2004). DoG is an approximation of LoG and is more computationally effective. In addition, Matas et al., proposed an affine invariant detector named Maximally Stable Extremal Regions (MSER) (Matas et al., 2004), which has been used for remote sensing image matching (Cheng et al., 2008). For the more comprehensive introduction to feature detectors, one may refer to the literature (Tuytelaars and Mikolajczyk, 2008; Mukherjee et al., 2015).

Once features such as interest points are detected, their descriptors for image matching must be constructed. A descriptor is often extracted based on a local region of an interest point. The extracted descriptors should be highly distinct and robust; in other words, they should adapt to various geometric and radiometric distortions. The most popular descriptor is the SIFT, which is a 3D histogram based on gradient magnitudes and orientations on a spatial arrangement consisting of a 4x4 square location grid. Due to its robustness to image rotation and scale changes, SIFT has been extensively applied to image matching. Inspired by the idea of SIFT, many researchers have proposed a variety of local descriptors to improve its capability. These descriptors mainly include the Gradient Location and Orientation Histogram (GLOH) (Mikolajczyk and Schmid, 2005), SURF, the Affine-SIFT (ASIFT) (Morel and Yu, 2009), the Center Symmetric Local Binary Pattern (CS-LBP) (Heikkila et al., 2009) and DAISY (Tola et al., 2010). GLOH is an extension of SIFT, which constructs the descriptor in a log-polar location grid instead of a square grid. The SURF descriptor is built by using the integral image technique and Haar wavelet responses, which is more computationally effective. The ASIFT descriptor simulates camera axis parameters to correct images and intends to address strong affine transformation. In the CS-LBP, a modified local binary pattern is integrated into the framework of SIFT for feature description. DAISY is an improved descriptor relative to SIFT and GLOH, which builds the histogram using a novel spatial arrangement composed of circular cells of varying sizes. Moreover, the DAISY-style arrangement has been shown to present the best performance among these various local descriptors (Kaneva et al., 2011; Winder et al., 2009). Recently, some binary descriptors have been developed for fast image matching, which include the binary robust independent elementary features (BRIEF) (Calonder et al., 2010) and the fast retina keypoint (FREAK) (Alahi et al., 2012). Furthermore, deep learning techniques also have been introduced for feature description to enhance the robustness (Simo-Serra et al., 2015; Zagoruyko and Komodakis, 2015). These methods mainly use the Convolutional Neural Networks to construct discriminant feature descriptors for image matching, and outperform the traditional descriptors such as SIFT and DAISY. However, they need a large number of training datasets, which may limit their broad use in remote sensing image matching because no public training datasets are available for that by now. In addition, these deep learning based methods are mainly designed to handle geometric distortions without fully considering complex radiometric changes between images.

In recent years, the remote sensing community has developed various automatic image matching methods by using local invariant features. Among these methods, SIFT is the most popular. Moreover, considering the characteristics of remote sensing images, researchers have proposed some improved SIFT algorithms such as the Uniform Robust SIFT (Sedaghat et al., 2011), the Adaptive Binning SIFT (Sedaghat and Ebadi, 2015b), and the Scale Restriction SIFT (Yi et al., 2008). These SIFT-based methods are mainly designed for addressing geometric distortions but are vulnerable to complex radiometric differences because the SIFT operator is based on local gradient