ISPRS Journal of Photogrammetry and Remote Sensing xxx (2018) xxx–xxx

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

A deep learning framework for remote sensing image registration ☆

Shuang Wang [a], Dou Quan [a], Xuefeng Liang [b,*], Mengdan Ning [a], Yanhe Guo [a], Licheng Jiao [a]

[a] Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, International Research Center for Intelligent Perception and Computation, Xidian University, Xi'an, Shaanxi Province 710071, China
[b] IST, Graduate School of Informatics, Kyoto University, Kyoto, Japan

ARTICLE INFO

Article history:
Received 30 May 2017
Received in revised form 1 December 2017
Accepted 26 December 2017
Available online xxxx

Keywords:
Deep neural network
Image registration
Remote sensing image
Self-learning
Transfer learning

ABSTRACT

We propose an effective deep neural network aiming at remote sensing image registration problem. Unlike conventional methods doing feature extraction and feature matching separately, we pair patches from sensed and reference images, and then learn the mapping directly between these patch-pairs and their matching labels for later registration. This end-to-end architecture allows us to optimize the whole processing (learning mapping function) through information feedback when training the network, which is lacking in conventional methods. In addition, to alleviate the small data issue of remote sensing images for training, our proposal introduces a self-learning by learning the mapping function using images and their transformed copies. Moreover, we apply a transfer learning to reduce the huge computation cost in the training stage. It does not only speed up our framework, but also get extra performance gains. The comprehensive experiments conducted on seven sets of remote sensing images, acquired by Radarsat, SPOT and Landsat, show that our proposal improves the registration accuracy up to 2.4–53.7%.

1. Introduction

Image registration is the process of geometrically aligning reference image and sensed image, which are about the same scene and acquired at different times, even by different sensors or from different viewpoints (Zitova and Flusser, 2003; Jacqueline, 2017). Image registration is a significant problem in remote sensing image processing, which will directly influence the performance of the follow-up works, such as image fusion, change detection, and environmental monitoring.

In past decade, the remote sensing image registration problem has been addressed by two types of methods: area-based methods and feature-based methods(Zitova and Flusser, 2003). Area-based methods search the optimal geometric transform parameters by optimizing the images similarity, in which Mutual Information (MI), Kullback-Leibler divergence, and normalized cross-correlation (NCC) are widely accepted (Kern and Pattichis, 2007; Suri and Reinartz, 2009; Parmehr et al., 2012; Liang et al., 2013; Xu et al., 2016b). Although area-based methods can be easily implemented, they are sensitive to intensity change, illumination change and noise. On the contrary, the feature-based methods overcome above defects and establish the geometric relation more effectively via matching the salient features, such as points, lines, and regions. In practice, SIFT (Lowe, 2004), SURF (Bay et al., 2008), HOG (Dalal and Triggs, 2005), MSER (Matas et al., 2004), Affine-SIFT (Morel and Yu, 2009) are commonly applied. Moreover, other researches combine area-based methods and feature-based methods for a coarse-to-fine image registration (Yong et al., 2009; Ma et al., 2010; Goncalves et al., 2011; Gong et al., 2014).

The representative in feature-based methods is scale-invariant feature transform (SIFT), because its feature descriptor is invariant under translation, rotation, and scale change on normal images. However, remote sensing images are generated by a complicated imaging mechanism, whose appearance is determined by the radiation characteristic, geometric characteristic of objects, and the transmitting or receiving configuration of sensors. In registration tasks, the reference image and sensed image may even come from different sensors, have varied resolutions, spectral and so on (Jacqueline, 2017). Due to the speciality of remote sensing images, the invariants of SIFT designed for normal images may not be maintained on remote sensing images. Our experiments reveal that the computed principal direction of SIFT keypoints becomes unreliable, because the statistic of gradients around the keypoint severely varies. An example, Fig. 1(a) and (b) illustrates this problem. We select some strong SIFT points that should be matched

☆ Special issue on Deep Learning for Remotely Sensed Data.

* Corresponding author.
E-mail addresses: shwang.xd@gmail.com (S. Wang), xliang@i.kyoto-u.ac.jp (X. Liang).

https://doi.org/10.1016/j.isprsjprs.2017.12.012
0924-2716/

Please cite this article in press as: Wang, S., et al. A deep learning framework for remote sensing image registration. ISPRS J. Photogram. Remote Sensing (2018), https://doi.org/10.1016/j.isprsjprs.2017.12.012

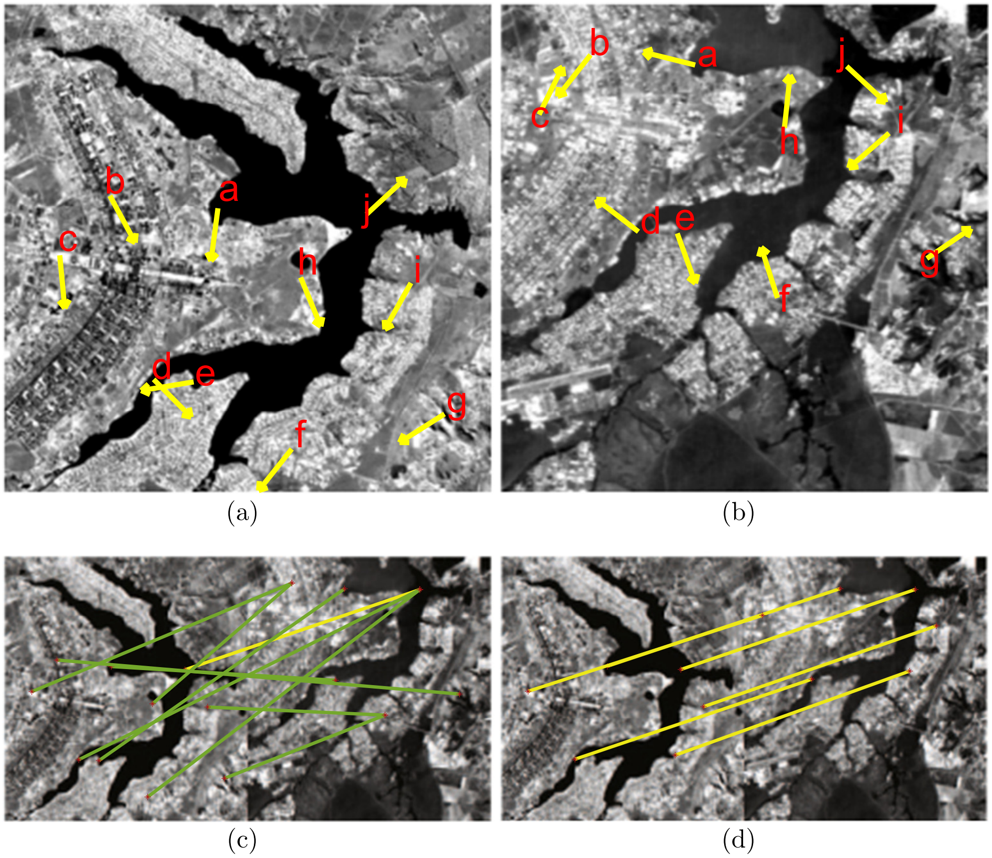

Fig. 1. (a) and (b) Show the corresponding keypoints between two remote sensing images, and their principle directions computed by SIFT algorithm. (c) and (d) Are matching results by SIFT and our method, respectively. The yellow lines are correct matching, whereas the green lines represent false matching.

between two images, and pair them by giving the same name. The yellow arrows indicate their principle directions in each of them. Obviously, the principal directions of corresponding points are not consistent or even going in the opposite direction. This unreliable direction results in a false matching as shown in Fig. 1(c). In this example, 91% SIFT points break down the rotation invariance, thus lead to a sever mis-registration due to insufficiency of correct matched points.

Moreover, the procedure of feature-based method is arguable for remote sensing image registration, which can be summarized as below. Firstly, detect keypoints from the reference image and sensed image, secondly, extract the features of these keypoints according to their neighborhood pixels, thirdly, match features by the feature distance, finally, the registration is done by estimating geometric transform matrix according to the matched keypoints. To ensure the performance, a well-designed feature and feature extractor are required, but need massive engineering works. These handcrafted features (e.g. edge, texture, corner and the statistical information of gradient) lack of high-level semantic information due to no information feedback between features extraction and matching. Therefore, these methods have limited applicability and solely perform well on specific images with suitable feature representation.

Recent years, deep learning has attracted increasing attention and achieved great successes. The major reason is that it is a fully data-driven scheme, can automatically learn the features from images. Specifically, deep learning has multi-level of non-linear operations, which seek to exploit the distribution structure of input data or abstract representation (Schmidhuber, 2014; Lecun et al., 2015). Its end-to-end architecture is able to optimize the entire network by the information feedback. Inspired by advantages of deep neural network (DNN), we propose a deep learning framework for remote sensing image registration. By taking the image patch-pairs as input and matching labels as output, our proposal directly learns an end-to-end mapping function. Its hidden layers correspond to features extractor, and the output layer corresponds to features matching. This architecture unifies the features extraction and matching in a closed-loop learning framework by involving not only information feedforward but also feedback. This significant difference, unlike the conventional feature-based method, permits the results of features matching to guide the process of features extraction, and then make the learned features more appropriate for the data. However, there are two encountered problems when we apply DNN: small training data set; and huge computation cost in training stage.

Deep learning is double-edged, can approximate very complex function but must optimize thousands (even millions) parameters for a problem. To prevent over-fitting, the networks are commonly trained on a large amount of samples. Unfortunately, it is tricky to have enough remote sensing images for this large scale data learning. Moreover, the manual annotation needs the professional knowledge, and the process is extreme expensive. In some exceptional cases, new sensed images, coming from a different sensor, are unlikely to be well registered by the network trained on data from other sensors. To address this problem, we proposed a self-learning with an idea of learning the mapping function from images and their varied transformed copies. This idea can be deemed to be a new way of data augmentation, which creates the labeled training data from scratch other than the conventional strategies by extending the dataset from the existing labeled data. This method comes with four advantages. First, the number of training samples are greatly increased. Second, the matching labels

Please cite this article in press as: Wang, S., et al. A deep learning framework for remote sensing image registration. ISPRS J. Photogram. Remote Sensing (2018), https://doi.org/10.1016/j.isprsjprs.2017.12.012