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

Pan-sharpening via deep metric learning

Yinghui Xing, Min Wang *, Shuyuan Yang, Licheng Jiao

School of Artificial Intelligence, Xidian University, Xi'an, Shaanxi Province 710071, China

ARTICLE INFO

Article history:
Received 31 May 2017
Received in revised form 21 January 2018
Accepted 23 January 2018
Available online xxxx

Keywords:
Pan-sharpening
Deep metric learning
Stacked Sparse AutoEncoders
Geometric multi-manifold embedding

ABSTRACT

Neighbors Embedding based pansharpening methods have received increasing interests in recent years. However, image patches do not strictly follow the similar structure in the shallow MultiSpectral (MS) and PANchromatic (PAN) image spaces, consequently leading to a bias to the pansharpening. In this paper, a new deep metric learning method is proposed to learn a refined geometric multi-manifold neighbor embedding, by exploring the hierarchical features of patches via multiple nonlinear deep neural networks. First of all, down-sampled PAN images from different satellites are divided into a large number of training image patches and are then grouped coarsely according to their shallow geometric structures. Afterwards, several Stacked Sparse AutoEncoders (SSAE) with similar structures are separately constructed and trained by these grouped patches. In the fusion, image patches of the source PAN image pass through the networks to extract features for formulating a deep distance metric and thus deriving their geometric labels. Then, patches with the same geometric labels are grouped to form geometric manifolds. Finally, the assumption that MS patches and PAN patches form the same geometric manifolds in two distinct spaces, is cast on geometric groups to formulate geometric multi-manifold embedding for estimating high resolution MS image patches. Some experiments are taken on datasets acquired by different satellites. The experimental results demonstrate that our proposed method can obtain better fusion results than its counterparts in terms of visual results and quantitative evaluations.

1. Introduction

In the optical imaging, it is sometimes difficult to obtain multispectral images with both high spectral and spatial resolutions due to the physical constraints on spatial and spectral resolutions (Alparone et al., 2016). Pansharpening aims at synthesizing a panchromatic image (who has high spatial resolution but low spectral resolution) and a multispectral image (with relatively high spectral resolution but low spatial resolution), to obtain High-Resolution MS (HRMS) images in both spectral and spatial domains, which are desirable for the subsequent tasks such as change detection (Gong et al., 2017a). During the last two decades, various pansharpening approaches have been proposed, which can typically be divided into three categories: Component-Substitution (CS) based methods, MultiResolution-Analysis (MRA) based methods and Model-based methods (Thomas et al., 2008). Among these three kinds of methods, CS- and MRA-based methods are classic methods that employ spectral transformation and spatial processing techniques respectively. The advantage of traditional CS-based methods is that they can produce good spatial-resolution enhanced images, however, they often result in severe spectral distortions. Based on this shortcoming, some adaptive methods are proposed, for example, Fast Intensity-Hue-Saturation transform with Spectral Adjustment (FIHS-SA) (Tu et al., 2004), adaptive Principal-Component-Aanalysis (Shah et al., 2008) and Adaptive Gram-Schmidt orthonormalization transform (GSA) (Aiazzi et al., 2007), and these methods can reduce color distortions effectively. In MRA-based methods, the high-frequency components are extracted by some MRA tools, such as Discrete Wavelet Transform (DWT) (Otazu et al., 2005), Support Value Transform (SVT) (Zheng et al., 2008) and Contourlet (Yang et al., 2010), and then injected into the resampled MS bands proportionally. The high-frequency information extracted from High Resolution Panchromatic (HRP) images should be adjusted before injected into the Low Resolution MultiSpectral (LRMS) images. As a result, some adjusting models include Inter-Band Structure Model (IBSM) (Thomas et al., 2008), Spectral Distortion Minimizing (SDM) (Garzelli and Nencini, 2005) injection model and Context-Based Decision (CBD) (Garzelli and Nencini, 2005) model are presented. Although the results of MRA-based methods have less

* Corresponding author.
E-mail addresses: yhxing_1@stu.xidian.edu.cn (Y. Xing), wangmin@xidian.edu.cn (M. Wang).

https://doi.org/10.1016/j.isprsjprs.2018.01.016
0924-2716/

Please cite this article in press as: Xing, Y., et al. Pan-sharpening via deep metric learning. ISPRS J. Photogram. Remote Sensing (2018), https://doi.org/10.1016/j.isprsjprs.2018.01.016

spectral distortion, they inevitably suffer from the aliasing and local dissimilarities. Due to this fact, some techniques are adopted, for example, as pointed out by Aiazzi et al. (2006), if the frequency response of the low-pass filter used in the multiscale decomposition matches the Modulation Transfer Function (MTF) of the spectral channel into which details are injected, the spatial enhancement of MRA-based methods is comparable to that of CS-based methods (Duran et al., 2017). Recently, there are some review papers that analyze the advantages and disadvantages of CS-based methods and MRA-based methods, and make a comparison between these two types of methods (Vivone et al., 2015; Pohl and Van Genderen, 1998; Aiazzi et al., 2017; Zhang, 2010; Zhang and Mishra, 2014).

The basic idea of the methods which belong to the third category is that there are some relationships among LRMS, HRP and HRMS images, so their goal is to explore such relationships to develop observation models from which HRMS can be derived. As Aiazzi et al. (2017) say, these methods perform optimizations according to some criteria, regardless of whether such an optimization is accomplished in the spectral or spatial domains. Intrinsically, they can be considered as regularized solutions of ill-posed problems, consisting of the reconstruction of the unknown high-resolution image from its coarse counterparts (Aly and Sharma, 2014). Among this category, various kinds of priors and models are utilized, for instance, the Bayesian framework has been used in a model-based fusion method, and an optional weighting parameter is added to the optimization model to balance the contribution between the spectral and spatial information preservation terms (Fasbender et al., 2008). Li and Leung (2009) using the Laplacian prior as a regularization term and the fusion task is thus modeled as a constrained deconvolution process, in which HRMS are unknown but a linear combination of them is available. Likewise, an AutoRegressive (AR) model is also employed as a regularization term to describe the spatial structure of HRMS images (Joshi et al., 2006). More recently, the maximum a posteriori (MAP) framework is utilized to describe the fusion problem (Zhang et al., 2012), where the edge-preserving Huber-Markov prior is employed to characterize the statistical properties of HRMS images.

Neighbors Embedding (NE) based pansharpening is a recently developed Model-based methods. Some of them generate High-Resolution (HR) patches via Locally Linear Embedding (LLE) (Roweis and Saul, 2000), which is a well-known manifold learning method whose goal is to find a low-dimensional embedding that best preserves the local geometry of data. Each datum is assumed to be linearly represented by its k nearest neighbors in a local region, and the low-dimensional embedding is calculated by the nearest neighbors and their weights. Liu et al. (2012) assumed that MS patches and corresponding PAN patches form the same manifolds and share similar intrinsic geometries although they are in two distinct spaces, then LLE is introduced to estimate MS patches by combining k candidate PAN patches selected from the dictionary. Otherwise, sparse representations (SR) have been used in many pansharpening methods by exploring the sparsity of source images to achieve better fusion results, which may also be treated as sparse neighbors embedding process. For example, a pansharpening method that is called "SparseFI" (Zhu and Bamler, 2013) constructs coupled dictionaries from HRP and its degraded version which has the same resolution as the LRMS, then the HRMS image is reconstructed by the HR dictionary and the corresponding sparse coefficients acquired by Low-Resolution (LR) dictionary. Based on the idea of coupled sparse representation technique in "SparseFI", some improved methods are proposed in succession, for example, Guo et al. (2014) put forward an Online Coupled Dictionary Learning (OCDL) approach in which a superposition strategy is applied to the dictionary construction, and Jiang et al. (2014) proposed a Two-Step Sparse Coding (TSSC) method to solve the problem appears when the structural information is weak in LRMS.

The preservation of local geometry of data in the embedding space is very critical in pansharpening. Most of available NE-based pansharpening methods believe that shallow features, such as first-order and second-order gradients, can better preserve the local geometry of PAN image patches. However, patches from real-world remote sending images are so diverse that patches will lie in multiple manifolds or subspaces of possibly different dimensions, and consequently manifolds may be very close to each other and have arbitrary dimensions and curvature (Elhamifar and Vidal, 2011). Therefore, high resolution PAN image patches do not strictly follow the similar structure in shallow feature space of low resolution PAN image space, which leads to a bias to the image fusion. On the other hand, it is well known that there are usually large variations in different patches of remote sensing images caused by numerous types of land covers and cluttered background, so the image patches in remote sensing images will lie in multiple nonlinear manifolds (Xiao et al., 2011), however, most of available NE-based pansharpening methods simply assume that different image patches define one manifold in the feature space regardless of how the patches are really distributed, which is not always valid to remote sensed data. In order to preserve local geometry in embedding space and accurately distinguish different manifolds in remote sensing images by hierarchical features rather than simply utilize the shallow features, an algorithm that has both strong discrimination ability and excellent geometric information preservation ability is desirable.

On the one hand, strong discrimination ability can give us a more accurate classification of different manifolds. On the other hand, the proper classification of different manifolds of patches in a remote sensed image is essential for the preservation of geometric information. Assume that an image patch is misclassified, then it may be reconstructed by a false manifold, which ultimately leads to the missing of correct geometric information about this patch. Distance metric techniques (Han et al., 2017; Yang and Jin, 2006) are effective to measure the differences among different manifolds that are important to determine which manifolds the new patches lie in. However, traditional distance metric algorithm usually learns a linear transformation to map samples into a new feature space, but remote sensed data have lower resolutions and cover more complicated land covers and backgrounds compared with natural images, which leads to the difficulties in capturing the nonlinear manifolds among remote sensed data. So it is insufficient for remote sensed data to utilize traditional distance metric strategies. In other words, a deep and hierarchical feature learning algorithm together with a metric learning strategy can be helpful for exploring and discriminating different manifolds in an image. Deep learning (DL) has been extensively used in machine learning and computer vision, and recently DL models also have extensively usage in remote sensing image processing field (Gong et al., 2017b; Cheng et al., 2016, 2015; Yao et al., 2016; Ma et al., 2016; Zhao and Du, 2016; Han et al., 2015). Due to its powerful nonlinearity representation ability, many researchers incorporated DL models into distance metric learning frameworks, i.e., Deep Metric Learning (DML). Deep Metric Learning has been widely used in the field of Computer Vision such as Face Verification (Hu et al., 2014) and Person Re-Identification (Yi et al., 2014), but it is seldom used in remote sensing image processing. In this paper, a new deep metric learning method is proposed to learn a refined geometric multi-manifold neighbor embedding, by exploring the hierarchical features of patches via multiple nonlinear deep neural networks. First of all, down-sampled PAN images from different satellites are divided into a large number of training image patches and are then grouped coarsely according to their shallow geometric structures. Afterwards, several Stacked Sparse AutoEncoders (SSAEs) with