Object-based Area-to-point Regression Kriging for Pansharpening

Yihang Zhang, Peter M. Atkinson, Feng Ling, Giles M. Foody, Qunming Wang, Yong Ge, Xiaodong Li, Yun Du

Abstract—**Optical Earth observation satellite sensors often provide a coarse spatial resolution (CR) multispectral (MS) image together with a fine spatial resolution (FR) panchromatic (PAN) image. Pansharpening is a technique applied to such satellite sensor images to generate a FR MS image by injecting spatial detail taken from the FR PAN image while simultaneously preserving the spectral information of MS image. Pansharpening methods are mostly applied on a per-pixel basis and use the PAN image to extract spatial detail. However, many land cover objects in FR satellite sensor images are not illustrated as independent pixels, but as many spatially aggregated pixels that contain important semantic information. In this paper, an object-based pansharpening approach, termed object-based area-to-point regression kriging (OATPRK), is proposed. OATPRK aims to fuse the MS and PAN images at the object-based scale and, thus, takes advantage of both the unified spectral information within the CR MS images and the spatial detail of the FR PAN image. OATPRK is composed of three stages: image segmentation, object-based regression and residual downscaling. Three datasets acquired from IKONOS and Worldview-2, and eleven benchmark pansharpening algorithms were used to provide a comprehensive assessment of the proposed OATPRK approach. In both the synthetic and real experiments, OATPRK produced the most superior pan-sharpened results in terms of visual and quantitative assessment. OATPRK is a new conceptual method that advances the pixel-level geostatistical pansharpening approach to the object-level, and provides more accurate pan-sharpened MS images.**

Index Terms—**Pansharpening, object-based, downscaling, geostatistics, segmentation, image fusion.**

Y. Zhang, F. Ling, X. Li, and Y. Du are with the Key Laboratory of Monitoring and Estimate for Environment and Disaster of Hubei Province, Innovation Academy for Precision Measurement Science and Technology, Chinese Academy of Sciences, Wuhan 430077, China;

P. M. Atkinson is with the Lancaster Environment Central, Faculty of Science and Technology, Lancaster University, Lancaster, LA1 4YQ, U.K;

G. M. Foody is with the school of Geography, University of Nottingham, University Park, Nottingham NG7 2RD, UK;

Q. Wang is with the College of Surveying and Geo-Informatics, Tongji University, Shanghai 200092, China;

Y. Ge is with the State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences & Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China.

I. INTRODUCTION

Remote sensing applications such as land cover/use change mapping, target recognition, hazard mapping, and environmental monitoring often require satellite sensor images with fine spectral, spatial, and temporal resolutions [1-7]. However, the expected design of such single satellite sensor system is limited by many competing constraints, and among the most important is the fundamental trade-off between the spatial and spectral resolutions of existing optical satellite sensor images [8, 9]. The spatial resolution of a satellite sensor image depends on the instantaneous field of view (IFOV) expressed as the ground area captured by one pixel; the smaller the IFOV, the finer the spatial resolution. Spectral resolution is regarded as the satellite sensor's signal electromagnetic bandwidth; the narrower the bandwidth, the finer the spectral resolution. To achieve a desired satellite image signal-to-noise ratio (SNR), a smaller IFOV should have relatively wider spectral bandwidth, and a narrower spectral bandwidth (more bands) should have a larger IFOV. Given such a trade-off between the IFOV and bandwidth, many optical satellite systems such as those associated with Landsat, SPOT, IKONOS, ALOS, GeoEye, Quickbird, OrbView and Worldview often have sensors with a set of coarse spatial resolution (CR) multispectral (MS) bands and a co-registered fine spatial resolution (FR) panchromatic (PAN) band. Given the widespread availability of CR MS coupled with FR PAN images, it is of great interest to produce both fine spatial and spectral resolution MS images by fusing the CR MS and FR PAN images, which is also a fusion goal termed pansharpening.

Pansharpening refers to the spatial sharpening of CR MS images by injecting spatial detail taken from a co-registered FR PAN image while simultaneously preserving the spectral information of the original CR MS images [10], and various pansharpening methods have been developed [9, 11]. One of the most classical types is the component substitution (CS) based pansharpening method, it projects the MS images into a new space and substitutes one component with the PAN image. Examples include principal component analysis (PCA) [12], intensity-hue-saturation (IHS) [13], Gram-Schmidt (GS) transform [14], Gram Schmidt Adaptive (GSA) [15], and partial replacement adaptive component substitution (PRACS) techniques [16]. However, if the wavelength range difference

between the MS and PAN images is too large, many CS-based methods may suffer from a significant spectral distortion during the inverse transformation [17], which limits severely their application to real situations. Another popular method is the multiresolution analysis (MRA)-based pansharpening which assumes that the high frequency information of the PAN image can be used to supplement the missing spatial detail within the CR MS images [11]. Compared with CS-based methods, MRA-based pansharpening algorithms, including the popular high-pass filtering (HPF) [18], smoothing filter-based intensity modulation (SFIM) [19], generalized laplacian pyramid (GLP) [20], à trous wavelet transform (ATWT) [21], additive wavelet luminance proportional (AWLP) [22] and wavelet-based fusion [23], usually deal with the image fusion process band-by-band, and it can prevent spectral distortion for the sharpened MS images [24]. With respect to the powerful performance of deep learning in remote sensing applications [25], Masi G. et al. developed the popular state-of-the-art pansharpening method based on convolutional neural networks (PNN) [26]. From then onwards, deep learning based on convolutional sparse representation [27], dynamic convolutional neural networks [28], deep residual networks [29, 30], bidirectional pyramid network [31] and interpretable deep network [32] were used for pansharpening. Especially, deep learning based pansharpening algorithms belong to a new generation of variational optimization based pansharpening method [33]. Moreover, approaches, for example, including regularized solutions of the inverse problem [34, 35], deconvolution-based filter estimation [36], morphological operators [37] also used to develop the variational optimization based pansharpening algorithms. Compared with CS- and MRA-based methods, the variational optimization based pansharpening method is expected to produce more accurate sharpened MS images [26].

As CR MS images are the only observed dataset, it is, therefore, important to make the sharpened MS images consistent to the observed CR MS images. Motivated by this, Pardo-Igúzquiza, Chica-Olmo and Atkinson [8] developed a geostatistical solution for image sharpening by considering the correlation and cross-correlation of CR MS images, in which downscaling cokriging (DSCK) is used to sharpen the CR MS images. DSCK and its updated spatial adaptive version [38] take the satellite sensor point spread function into account and have the advantages of preserving the spectral information of the observed CR MS images. Atkinson, Pardo-Iguzquiza and Chica-Olmo [39] further improved the DSCK to produce sharpened MS images that have finer spatial resolution than any of the input MS and PAN images. Thereafter, the kriging with external drift (KED) model [40], which requires only the estimation of direct autovariograms or variograms was proposed for MODIS (MODerate-resolution Imaging Spectroradiometer) image downscaling, and it is easier to implement and render operational than DSCK. In general, both

DSCK and KED are one-stage models which exploit the spatial details of the PAN image and preserve the spectral information of the CR MS images at the same time. Although these one-stage models are mathematically perfect, it may suffer from expensive computational cost and cannot make full use of the spatial detail in the PAN image. Area-to-point regression kriging (ATPRK) was then proposed for downscaling MODIS images [41]. Compared with DSCK and KED, ATPRK is more efficient and, like DSCK, can preserve perfectly the spectral information of the original CR MS images, in which the first stage of ATPRK is linear regression modeling and the second stage is residual downscaling. ATPRK was further improved by making it spatially adaptive (AATPRK) [42]. ATPRK and AATPRK were proved to produce more accurate results than most CS- and MRA-based pansharpening algorithms [43].

Pansharpening is generally implemented through per-pixel based spectral-spatial fusion, since almost all techniques extract spatial detail in PAN image at the per-pixel scale [44]. In real applications, it is reasonable to assume that spatial detail in the PAN image varies from pixel-to-pixel for CR satellite sensor images, such as from MODIS, MERIS and the most recently launched Sentinel-3 satellite, because these images contain pixels covering a large area (e.g. 500 m × 500 m or 300 m × 300 m). As the increasing of the spatial resolution of satellite images, the pixels would represent smaller areas, many objects in images are likely to change from being of sub-pixel size to being comprised of a region of pixels that represent important semantic information [45, 46]. For FR satellite sensor image classification, the above per-pixel based methods have been shown to have a negative influence on the results [47-49]. It follows that such a phenomenon may also influence the results of pansharpening in supplying spatial detail for sharpened images from the PAN image [50, 51].

Object-based classification methods have been developed widely to increase the accuracy of classification of remotely sensed images at the pixel scale [45, 48, 52, 53] and even the sub-pixel scale [54, 55]. Object-based analysis has a long history with much early work using object boundaries defined in ancillary data such as cadastral and topographic maps to provide meaningful spatial units [56], but are now generated commonly from FR data via segmentation analysis [57]. Segmentation aims to cluster an image into spatially continuous and homogeneous groups, which are regarded as the "objects" for further processing through object-based classification. Over the past two decades, numerous segmentation algorithms have been proposed for remote sensing applications [58, 59, 48], and fuzzy models are the most widely used due to their ability to handle complexity and their high efficiency [60]. Irrespective of how the objects are defined, object-based classification methods consider objects rather than pixels as the basic spatial unit and classify each of the individual objects into different land cover types. This significant change of the spatial unit for classification from image pixel to land cover object is a
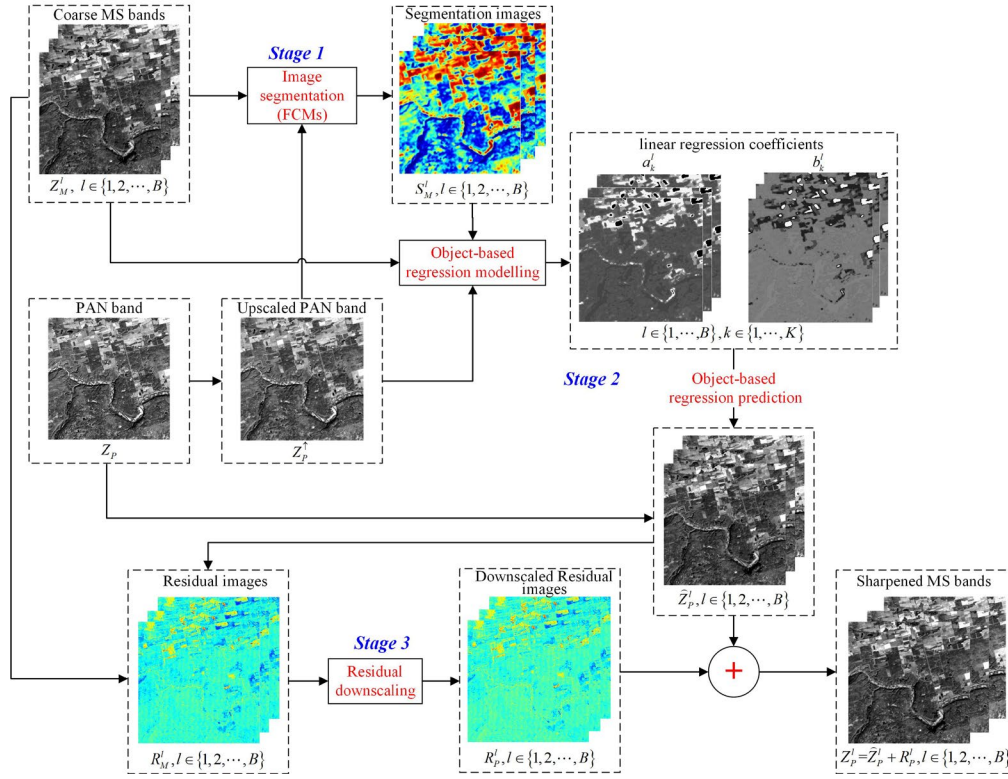
Fig. 1. Flowchart of the proposed OATPRK.

milestone in the development of satellite sensor image classification, because it reduces the within-class spectral variation and removes the so-called salt-and-pepper effect to some extent [61], as well as providing a more natural spatial unit for some phenomena.

Motivated by the development of object-based classification, this research aims to propose an object-based pansharpening approach to advance on current per-pixel level geostatistical pansharpening method and, thereby, further increase the accuracy of the sharpened MS. It is noted that there are a few kinds of researches using segmentation image as auxiliary data to improve the pansharpening results, such as the GLP and GSA based on the binary partition tree (GLP-BPT and GSA-BPT) [62], GLP based on the shuffled complex evolution [63], GLP based on robust regression [64]; but unfortunately, the key concept of 'object' was not yet introduced [65, 62, 66]. The proposed object-based pansharpening approach, termed object-based area-to-point regression kriging (OATPRK), holds the primary objective of exploiting more accurate spatial detail from the FR PAN image by using the new concept of object units for geostatistical pansharpening method, and perfectly preserving the spectral information of the original CR MS images at the same time. OATPRK is implemented band-by-band, and it is composed of three main stages: image segmentation, object-based regression and residual downscaling. In the image segmentation part, each band of the CR MS images and the upscaled PAN image are integrated to generate the synthetic two-band CR MS images. Hereafter, the spatial constraints based fuzzy c-means (FCM_S) [67]

segmentation algorithm is applied for the synthetic two-band MS images to produce segmented images. In the object-based regression part, the relationship between each band of the CR MS images and upscaled PAN image is estimated for each object that is labelled by the segmentation images, and then they are used to predict the initial sharpened MS images. Finally, the residual downscaling part is used to preserve the spectral information in the initial sharpened MS images of the object-based regression modelling.

Compared with existing pansharpening approaches, OATPRK holds the following desirable properties: 1) OATPRK overcomes the traditional per-pixel based perspectives in geostatistical pansharpening method, and provides a simple, yet efficient framework of object-based pansharpening; 2) OATPRK not only takes advantage of the spatial detail of the PAN image, but also the spectral information within the upscaled PAN image to generate a segmentation image for each band of the CR MS image. In this way, more accurate spatial detail in the PAN image can be exploited by OATPRK; 3) OATPRK applies a spatial constraints-based clustering algorithm to produce more robust segmentation images in the image segmentation part; 4) OATPRK inherits the superiority of perfectly preserving the spectral information of the observed CR MS images in the sharpened MS images.

The remainder of this paper is organised as follows: The principle and method of OATPRK are introduced in Section II. Section III presents results of both synthetic and real data sets,

and section IV provides a discussion of the performance of OATPRK. Conclusions are summarized in Section V.

II. METHOD

Suppose that Z_M^l is the lth band of the CR MS images (l=1, 2, …, B, where B is the band number), Z_P is the co-registered FR PAN image, and F is the spatial ratio between Z_M^l and Z_P. Subscripts M and P in Z_M^l and Z_P indicate the MS images and PAN image, respectively. OATPRK aims to predict the sharpened FR Z_P^l from each CR image Z_M^l with the help of Z_P. Fig. 1 shows a flowchart describing the methodology of OATPRK, including three main stages: image segmentation with FCM_S, object-based regression, and residual downscaling. More details are given in the following sections.

A. Image segmentation with FCM_S

Let Z_{MP}^l be the synthetic two-band CR MS images, where l is set be 1 or 2. Z_{MP}^l is composed of each band of the CR MS image and the upscaled PAN band $Z_P^{\uparrow_F}$, which is the upscaled version of PAN band Z_P with spatial ratio of F and $\uparrow_F$ denotes an upscaling operation. Specially, the calculation of CR PAN band $Z_P^{\uparrow_F}$ is expressed as

$$Z_P^{\uparrow_F} = H(\mathbf{x}) * Z_P(\mathbf{x}) = \int H(\mathbf{x}-\mathbf{y})Z_P(\mathbf{y})d\mathbf{y} , \quad (1)$$

where $\mathbf{x}$ and $\mathbf{y}$ mean all of the coarse pixels in each band of the CR MS images Z_M^l, $H(\mathbf{x})$ denotes the point spread function between CR $Z_P^{\uparrow_F}$ and FR Z_P, and $*$ is the convolution operator. Therefore, Z_{MP}^l can be considered as a combination of Z_M^l and upscaled PAN image $Z_P^{\uparrow_F}$, and is expressed as $Z_{MP}^l = [Z_M^l, Z_P^{\uparrow_F}]$.

With the input dataset of Z_{MP}^l, FCM_S algorithm is used here to convert it into a segmentation image, due to its superiority in corrupting with salt-and-pepper noise by taking account of spatial contextual information [68-70]. FCM_S aims to cluster the dataset into K clusters by minimizing the objective function given by

$$J_{\min} = \sum_{k=1}^{K}\sum_{i=1}^{N}\sum_{l=1}^{2} u_{ik}^m \left\| r(\mathbf{x}_i,l)-v_k(l)\right\|^2 + \\ \alpha\sum_{k=1}^{K}\sum_{i=1}^{N}\sum_{l=1}^{2} u_{ik}^m \left\| \overline{r}(\mathbf{x}_i,l)-v_k(l)\right\|^2 , \quad (2)$$

where $\mathbf{x}_i$ is the spatial location of a coarse pixel i in each band of Z_{MP}^l (i=1, 2, …, N, and N is the number of coarse pixels in each band of Z_{MP}^l), $r(\mathbf{x}_i,b)$ is the spectrum of the coarse pixel in the lth band located at $\mathbf{x}_i$, $v_k(l)$ is the cluster center value of the kth cluster in the lth band, u_{ik}^m indicates the fuzzy membership value of coarse pixel $\mathbf{x}_i$ belonging to the kth cluster, m is a weighting exponent of fuzzy membership. $\overline{r}(\mathbf{x}_i,l)$ is the mean spectrum of $w \times w$ neighboring pixels centered at $\mathbf{x}_i$ in the lth band (w is the size of the local window), and α is a parameter used to control the contribution of the spatial constraints in the objective function. Similarly to standard FCM [71], minimization of the objective

function $J_{\min}$ can be obtained by updating the fuzzy membership u_{ik}^m and clustering centers $v_k(l)$ with the following equations

$$u_{ik}^m = \frac{\left(\left\| r(\mathbf{x}_i,b)-v_k(l)\right\|^2 + \alpha\left\|\overline{r}(\mathbf{x}_i,l)-v_k(l)\right\|^2\right)^{\frac{-1}{m-1}}}{\sum_{j=1}^{K}\left(\left\| r(\mathbf{x}_i,b)-v_j(l)\right\|^2 + \alpha\left\|\overline{r}(\mathbf{x}_i,l)-v_j(l)\right\|^2\right)^{\frac{-1}{m-1}}}, \quad (3)$$

$$v_k(l) = \frac{\sum_{i=1}^{N} u_{ik}^m(r(\mathbf{x}_i,b)+\alpha\overline{r}(\mathbf{x}_i,l))}{(1+\alpha)\sum_{i=1}^{N} u_{ik}^m}. \quad (4)$$

Suppose that S_M^l is the segmentation image with K clusters. With equations (2), (3), and (4), the segmentation image S_M^l with K clusters can be generated from the synthetic CR MS images Z_{MP}^l (l=1 and 2). More information about the implementation of the FCM_S image segmentation process is presented in Chen and Zhang [68]. It is noted that there are two other ways to complete the process of image segmentation: 1) use the upscaled PAN band as input; 2) use each band of CR MS images as input. But there is phenomenon that objects in different spectral bands will present various spatial patterns. If the segmentation is focused only on the CR MS band, it cannot match the objects in upscaled PAN band; otherwise, the segmentation result cannot match the objects in CR MS band if only focusing on the upscaled PAN band. Therefore, using both the upscaled PAN band and each CR MS band as the input dataset to produce the unique segmentation image is a much better trade-off choice.

B. Object-based regression

Object-based regression is used to inject spatial detail from the PAN image Z_P into the sharpened image $\hat{Z}_P^l$ for each input CR image Z_M^l based on the units of the object (cluster). It is assumed that the linear relationship between the observed CR image Z_M^l and upscaled PAN image $Z_P^{\uparrow_F}$ varies from object to object, and pixels belonging to the same object in the segmentation image S_M^l share the same linear relationship. As shown in Fig. 1, object-based regression is composed of two parts, object-based regression modelling and object-based regression prediction:

1) *Object-based regression modelling*: The relationship between the objects of the observed CR image Z_M^l and upscaled PAN image $Z_P^{\uparrow_F}$ are modelled as a linear regression function

$$Z_P^{\uparrow_F}(\mathbf{X}_k) = a_l(\mathbf{X}_k)Z_M^l(\mathbf{X}_k) + b_l(\mathbf{X}_k) + R_l(\mathbf{X}_k), \forall k=1,\cdots,K \quad (5)$$

in which $\mathbf{X}_k$ indicates the locations of all coarse pixels belonging to the kth cluster (object) in the segmentation image S_M^l, $R_l(\mathbf{X}_k)$ is the residual term of the linear regression model for the kth cluster, $a_l(\mathbf{X}_k)$ and $b_l(\mathbf{X}_k)$ are two linear regression coefficients for the kth cluster. Since Z_M^l and $Z_P^{\uparrow_F}$ are already known, and S_M^l can provide the pixel locations of different objects, ordinary least squares can be used to estimate $a_l(\mathbf{X}_k)$ and $b_l(\mathbf{X}_k)$ for different objects of the lth band. As shown in Fig. 2, if
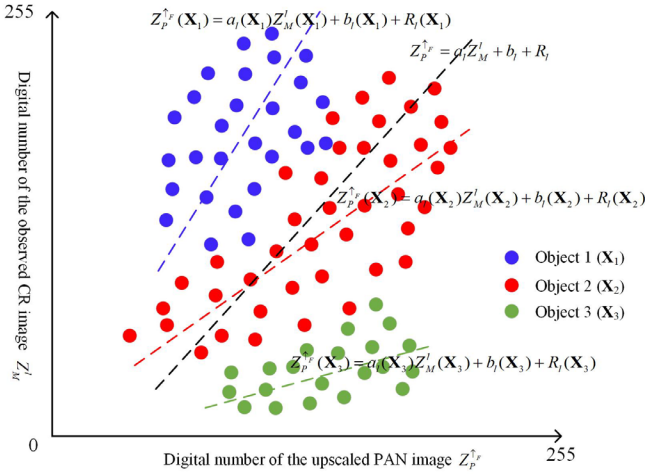
Fig. 2. A scatterplot used to illustrate the relationship between different objects of the observed CR image Z_M^l and upscaled PAN image $Z_P^{\uparrow_F}$.

all of the pixels are regarded as one object, the relationship between the observed CR image Z_M^l and the upscaled PAN image $Z_P^{\uparrow_F}$ is reprsented as the black line; but in fact, it cannot accurately reflect the relationships between the three objects (blue, red, and green points) of the observed CR image and upscaled PAN image, especially for the blue and green objects.

2) *Object-based regression prediction*: It is assumed that the relationship of different objects fitted between the CR image Z_M^l and upscaled PAN image $Z_P^{\uparrow_F}$ with equation (5) is invariant between the predicted sharpened FR image $\hat{Z}_P^l$ and FR PAN image Z_P. The intermediate sharpened image $\hat{Z}_P^l$ is, therefore, calculated as

$$\hat{Z}_P^l(\mathbf{X}_k) = a_l(\mathbf{X}_k)Z_P(\mathbf{X}_k)+b_l(\mathbf{X}_k), \quad \forall k=1,2,\cdots,K. \quad (6)$$

It is noted that for the regression part of OATPRK, the key task is to inherit as much spatial detail information in the FR PAN image as possible. But many spatial information in the PAN image will be overfitted if a nonlinear regression model is applied for the PAN image. Therefore, linear regression model is used here, to preserve full spatial information of the PAN image. Although using a linear regression model would lead to serious spectral distortion, it will be perfectly solved in the following residual downscaling process [72, 64].

C. Residual downscaling

In the object-based regression process, the FR image $\hat{Z}_P^l$ is predicted from the observed CR image Z_M^l. Ideally, $\hat{Z}_P^l$ should be strictly consistent to Z_M^l, but there is no constraint in the regression process to ensure this, and residuals between $\hat{Z}_P^l$ and Z_M^l exist inevitably. Let $\hat{Z}_{PM}^l$ be the upscaled image of the predicted $\hat{Z}_P^l$ produced by using the upscaling operation shown in equation (1). The coarse residual image between $\hat{Z}_{PM}^l$ and the observed Z_M^l is expressed as Z_{MR}^l given by

$$Z_{MR}^l(\mathbf{x}_i) = Z_M^l(\mathbf{x}_i) - \hat{Z}_{PM}^l(\mathbf{x}_i). \quad (7)$$

To make the predicted FR image consistent to the observed CR MS image and take advantage of the spectral information within the observed CR image, the coarse residual image Z_{MR}^l

should be spatially downscaled to the target spatial resolution of the PAN image and supplemented to predict the FR MS image. Let Z_{PR}^l be the downscaled FR residual image generated from the CR residual image Z_{MR}^l with a spatial downscaling operation $\downarrow_F$, and it is expressed as

$$Z_{PR}^l = Z_{MR}^l \downarrow_F =[Z_M^l(\mathbf{x}_i) - \hat{Z}_{PM}^l(\mathbf{x}_i)]\downarrow_F. \quad (8)$$

The spatial downscaling operation $\downarrow_F$ can be implemented by simple spatial interpolation approaches, such as spline, bilinear and bicubic interpolation, but these algorithms cannot increase the information content of the downscaled image and additional residuals could also be introduced into the downscaled images of these ordinary spatial interpolation algorithms [73]. By contrast, area-to-point kriging (ATPK) is a geostatistical-based downscaling method, which treats each observed datum as a centroid by taking account of the spatially surrounding coarse pixels and size of support [74, 75], and holds the key superiority of coherence between the observed data and predicted data [76]. Therefore, ATPK was chosen as the spatial downscaling operation $\downarrow_F$.

Based on ATPK, the fine pixel value $Z_{PR}^l(\mathbf{x}_j)$ with a spatial location of $\mathbf{x}_j$ is estimated as a spatially weighted combination of the observed coarse pixels, and is expressed as

$$Z_{PR}^l(\mathbf{x}_0)=\sum_{i=1}^{N_0} \lambda_i Z_{MR}^l(\mathbf{x}_i), \quad (9)$$

$$\text{Subject to:} \sum_{i=1}^{N_0} \lambda_i =1, \quad (10)$$

where N_0 is the number of coarse pixels in the neighboring system (e.g. 5×5 or 7×7 window of coarse pixels centered at fine pixel located at $\mathbf{x}_0$), λ_i is the weight of coarse pixel in the neighboring system. As $Z_{MR}^l(\mathbf{x}_i)$ is already known, the estimation of $Z_{PR}^l(\mathbf{x}_j)$ requires the estimation of λ_i, and this can be solved by the following kriging system

$$\sum_{i=1}^{N_0} \lambda_i \varphi_{MM}^l(\mathbf{x}_i,\mathbf{x}_j)+\theta_0 = \sum_{i=1}^{N_0} \varphi_{PM}^l(\mathbf{x}_i,\mathbf{x}_0), \quad (11)$$

$$\varphi_{MM}^l(\mathbf{x}_i,\mathbf{x}_j) = \varphi_{PP}^l(\mathbf{h}_{Tt}) * \psi_M^l(\mathbf{h}_{Tt}), \quad (12)$$

$$\varphi_{PM}^l(\mathbf{x}_i,\mathbf{x}_0) = \varphi_{PP}^l(\mathbf{h}_T) * \psi_M^l(\mathbf{h}_T), \quad (13)$$

in which θ_0 is the Lagrange multiplier, and $*$ is the convolution operator, $\varphi_{MM}^l(\mathbf{x}_i,\mathbf{x}_j)$ is the coarse-to-coarse semivariogram calculated by the point support covariance between coarse pixels located at $\mathbf{x}_i$ and $\mathbf{x}_j$ as shown in equation (12), $\varphi_{PM}^l(\mathbf{x}_i,\mathbf{x}_0)$ is the fine-to-coarse semivariogram calculated by the point support covariance between fine pixel located at $\mathbf{x}_0$ and coarse pixel located at $\mathbf{x}_i$ as shown in equation (13), $\mathbf{h}_{Tt}$ denotes the Euclidean distance between fine pixels in coarse pixels located at $\mathbf{x}_i$ and $\mathbf{x}_j$, $\mathbf{h}_T$ is the Euclidean distance between fine pixel located at $\mathbf{x}_0$ and fine pixel in coarse pixel located at $\mathbf{x}_j$. $\varphi_{PP}^l(\mathbf{h})$ is the fine-to-fine semivariogram calculated by the deconvolution of the coarse semivariogram $\varphi_M^l(\mathbf{h})$, which is expressed as

$$\varphi_M^l(\mathbf{h})=\frac{1}{2N_\mathbf{h}}\sum_{i=1}^{N_\mathbf{h}} \left\| Z_{MR}^l(\mathbf{x}_i) - Z_{MR}^l(\mathbf{x}_i + \mathbf{h}) \right\|^2, \quad (14)$$
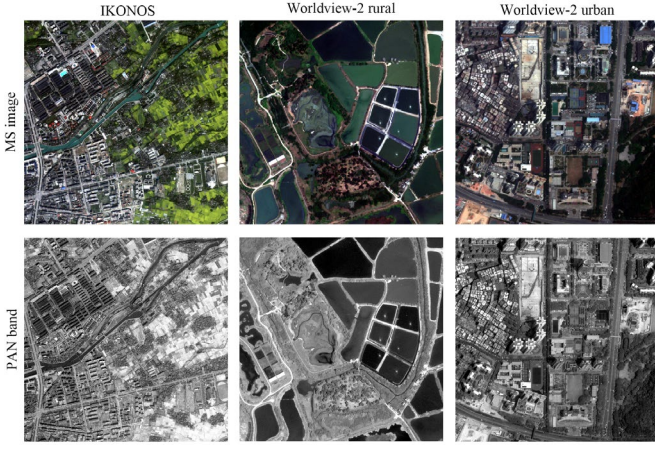
Fig. 3. Three datasets including IKONOS (400 × 400 pixels), Worldview-2 rural (400 × 400 pixels) and Worldview-2 urban (400 × 400 pixels) MS images and corresponding PAN band (1600 × 1600 pixels) used in the synthetic (reduced resolution) and real (full resolution) experiments.

TABLE I
BANDWIDTH AND SPATIAL RESOLUTION OF THE MS AND PAN BANDS FOR THE TWO USED DATASETS.

Dataset	MS		PAN	
	Bandwidth (nm)	SSI(m)	Bandwidth (nm)	SSI(m)
IKONOS	Blue (450-520); Green (510-600); Red (630~700); NIR (760~850)	4	450~900	1
Worldview-2	Coastal (400-450); Blue (450-510); Green (510-580); Yellow (585-625); Red (630~690); Red edge (705-745); NIR (770~895); NIR2 (860-1040)	2	450~800	0.5

Note: SSI means spatial sampling interval.

where $N(\mathbf{h})$ is the number of paired coarse pixels at lag vector $\mathbf{h}$ centered at the coarse pixel located at $\mathbf{X}_i$. More details about the point support semivariogram used in deconvolution can be found in [41]. $\psi_M^l(\mathbf{h})$ is the point spread function (PSF) of the satellite sensor, and it is assumed to be the Gaussian filter in the following equation

$$h_M(\mathbf{x}) = \begin{cases} \dfrac{1}{2\pi\sigma^2}\exp\left[-\left(\dfrac{x_1^2+x_2^2}{2\sigma^2}\right)\right], & \text{if } \mathbf{x} \in M(\mathbf{x}) \\ 0, & \text{otherwise} \end{cases}, \quad (15)$$

in which σ is the width of Gaussian PSF, the coordinates of location $\mathbf{x}$ is expressed as $\{x_1, x_2\}$, and $M(\mathbf{x})$ is the neighborhood system of the pixel centered at $\mathbf{x}$. Moreover, all of the above upscaling operation of $\uparrow_F$ in the above three stages is based on the PSF in equation (15).

D. *OATPRK*

The predicted FR MS image Z_P^l of OATPRK is estimated as the combination of the intermediate sharpened image $\hat{Z}_P^l$ generated in the above object-based regression part and the above downscaled residual image $Z_{PR}^l(\mathbf{x}_j)$ generated in the residual downscaling part. Therefore, Z_P^l is expressed as

$$Z_P^l(\mathbf{x}_j) = \hat{Z}_P^l(\mathbf{x}_j) + Z_{PR}^l(\mathbf{x}_j), \quad (16)$$

in which $\mathbf{x}_j$ is the spatial location of a fine pixel j at the spatial resolution of the PAN image (j=1, 2, …, $N\times F^2$). The detailed implementation of OATPRK for fusing the CR MS images and FR PAN image is presented in the flowchart of Fig. 1. It is noteworthy that OATPRK is performed band-by-band, which is similar to the MRA-based pansharpening methods. Once all of the bands of the CR MS image are processed, in turn, OATPRK produces FR MS images that have the same spatial resolution as that of the PAN image and have the same band number as that of the CR MS images.

For OATPRK, the three main stages have different objectives to make the best use of both the CR MS images and FR PAN image. Particularly, image segmentation aims to take advantage of the integrated spectral information of the CR MS images and PAN image, and produce object units that have similar spectral characteristics between each of the CR MS images and upscaled PAN image. These object units are then used in the second object-based regression part to exploit more accurate spatial detail of various objects from the PAN image for each sharpened FR MS image. With traditional perspectives, the PAN image can only be used to provide spatial detail (high frequency) in the pansharpening process. In fact, spatial detail of different objects in the PAN image, such as roads, open water, vegetation, bare soil, buildings, and so on, share different spectral responses in different wavelength bands. If we exploit the spatial detail of different objects for all of the CR MS image bands based on the whole fine pixels of the PAN image in the regression part, the spatial detail for some objects may not be enhanced and many unwanted spatial details of other objects may be over-enhanced, because all of the pixels share the same or spatially local linear regression coefficients a_l and b_l in the regression prediction. By contrast, if we take into account the spectral information of the PAN image and fit unique regression models for different objects as done in the object-based regression part of OATPRK, the spatial detail of different objects can, thus, be accurately enhanced in the resultant sharpened MS images. Therefore, the superiority of OATPRK is mainly coming from the object-based regression part, as it can exploit more accurate spatial information from the FR PAN band for each band of the CR MS images. In the residual downscaling part, OATPRK is based on ATPK and inherits the superiority of perfectly preserving the spectral information of observed CR MS images in the sharpened MS images.

III. EXPERIMENTAL RESULTS

A. *Datasets*

In this section, experiments were carried out on many optical satellite sensors acquiring MS images together with a PAN band to provide a comprehensive assessment of the performance of OATPRK. With respect to the quantitative validation of OATPRK, both synthetic (reduced resolution) and

TABLE II
QUANTITATIVE ASSESSMENT OF THE PAN-SHARPENING RESULTS FOR THE IKONOS AND WORLDVIEW-2 DATASETS IN THE SYNTHETIC EXPERIMENT.

	IKONOS				Worldview-2 rural				Worldview-2 urban			
	UIQI	ERGAS	SAM	Q4	UIQI	ERGAS	SAM	Q8	UIQI	ERGAS	SAM	Q8
Ideal	1	0	0	1	1	0	0	1	1	0	0	1
GSA	0.8509	2.0887	2.9054	0.8362	0.9308	1.7167	1.9698	0.8694	0.9721	1.3879	2.1697	0.9394
BDSD	0.9165	1.7926	2.5347	0.8683	0.9481	1.6701	2.2006	0.8626	0.9671	1.4366	2.2668	0.9369
PRACS	0.8242	2.2384	2.9458	0.808	0.8947	1.8207	2.0321	0.8160	0.9499	1.7089	2.3776	0.9158
HPF	0.8749	2.0546	2.7561	0.8279	0.9318	1.6920	1.9927	0.8617	0.9523	1.6079	2.1657	0.9150
AWLP	0.8865	1.9572	2.6625	0.8468	0.9331	1.8376	2.0594	0.8661	0.9585	1.5578	2.2108	0.9272
MF-HG	0.9004	1.8875	2.6495	0.8558	0.9454	1.6257	1.9236	0.8694	0.9653	1.4329	2.1105	0.9347
FE	0.8891	1.9390	2.6310	0.8373	0.9395	1.5490	1.8693	0.8691	0.9585	1.5283	2.1603	0.9222
PNN	0.8898	2.3362	3.5682	0.7959	0.9392	3.0808	6.2134	0.4171	0.9655	2.4899	5.2066	0.8155
GLP-BPT	0.8923	1.9146	2.6688	0.8464	0.9447	1.8532	2.2156	0.8540	0.9695	1.3941	2.0310	0.9323
GSA-BPT	0.8975	1.8616	2.6588	0.8534	0.9444	1.8336	2.2250	0.8490	0.9702	1.3030	1.9942	0.9396
ATPRK	0.8826	1.9386	2.6437	0.8574	0.9493	1.4933	1.7549	0.8884	0.9746	1.2623	1.9395	0.9485
OATPRK	**0.9404**	**1.5759**	**2.2117**	**0.8936**	**0.9684**	**1.2785**	**1.6063**	**0.9008**	**0.9770**	**1.1883**	**1.8038**	**0.9495**

(a) 16 m degraded MS image   (b) 4 m degraded PAN band   (c) 4 m reference MS image   (d) Segmentation image (band 1)

(e) GSA   (f) BDSD   (g) PRACS   (h) HPF

(i) AWLP   (j) MF-HG   (k) FE   (l) PNN

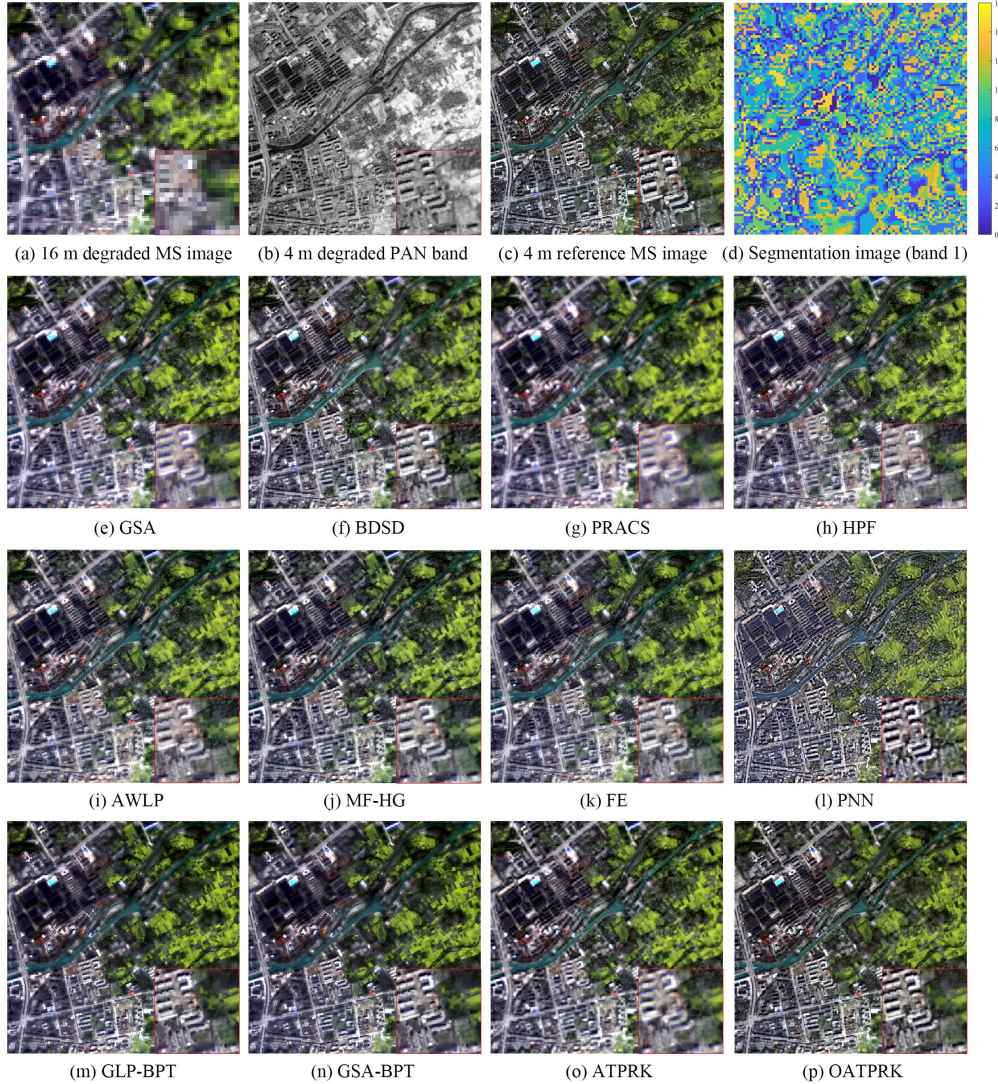(m) GLP-BPT   (n) GSA-BPT   (o) ATPRK   (p) OATPRK

Fig. 4. Results of different pansharpening algorithms for the IKONOS dataset (RGB: bands 321).

real (full resolution) image experiments were designed. Specifically, three datasets, including one IKONOS image and two Worldview-2 images as shown in Fig. 3, were used, and all of them were used for the synthetic and real experiments. Table I lists some key features of the datasets, and details about them are presented as the following subsections.

1) IKONOS dataset: This dataset acquired by the IKONOS sensor covers a suburb area of Sichuan in China. The

IKONOS sensor's MS image has four bands including the blue, green, red, and near-infrared wavelengths and a PAN band channel. The spatial resolution of the IKONOS MS images is 4 m and that of the PAN band is 1 m. The spatial size of the used MS image is 400 × 400 pixels, and the corresponding PAN image is 1600 × 1600 pixels.

2) Worldview-2 rural and urban datasets: These two datasets acquired by the Worldview-2 sensor represent a rural

(a) 8 m degraded MS image　　(b) 2 m degraded PAN band　　(c) 2 m reference MS image　　(d) Segmentation image (band 1)

(e) GSA　　(f) BDSD　　(g) PRACS　　(h) HPF

(i) AWLP　　(j) MF-HG　　(k) FE　　(l) PNN

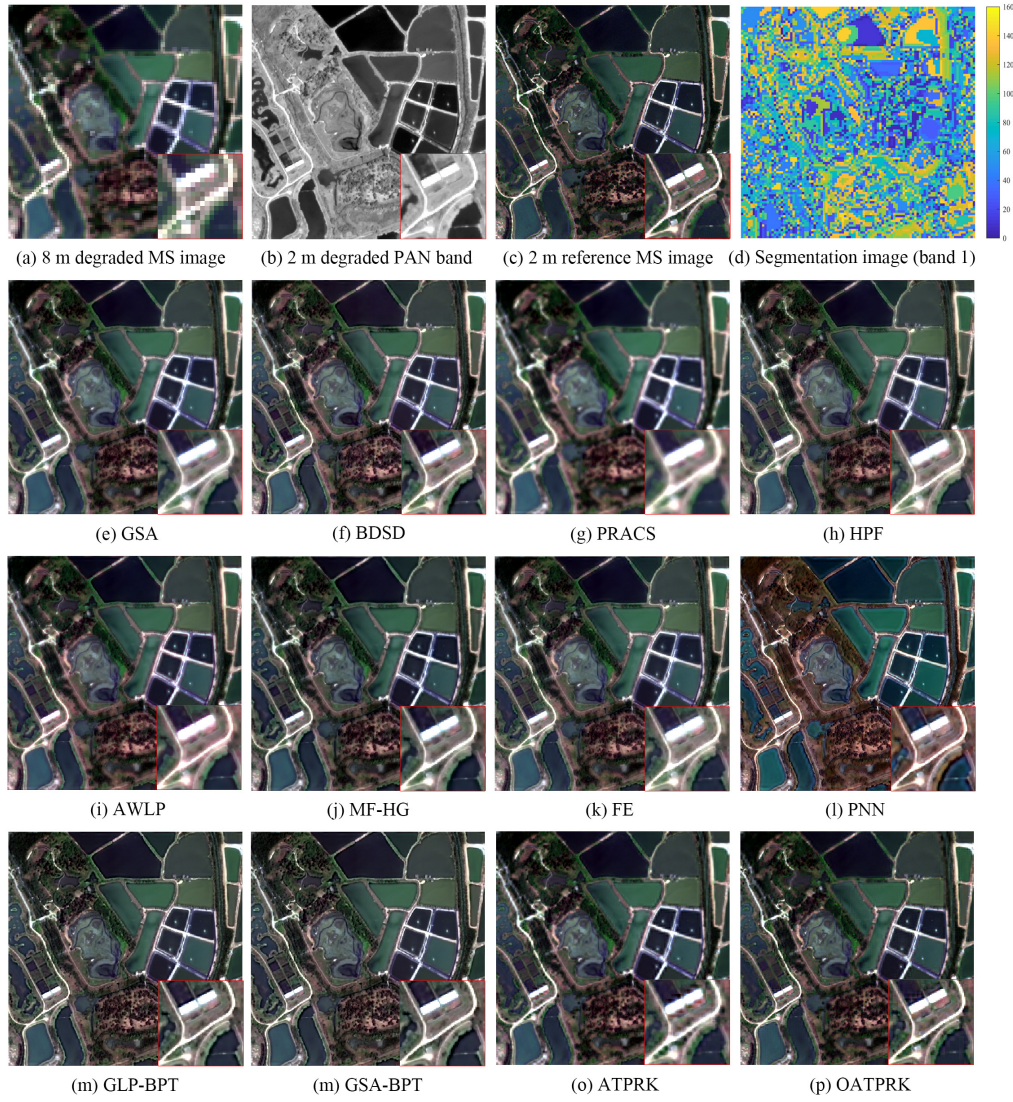(m) GLP-BPT　　(m) GSA-BPT　　(o) ATPRK　　(p) OATPRK

Fig. 5. Results of different pansharpening algorithms for the Worldview-2 rural dataset (RGB: bands 532).

area of Hong Kong and an urban area of Shenzhen in China. The MS image of Worldview-2 contains eight bands covering four standard (red, blue, green, near-infrared) bands and four new (red edge, coastal, yellow, near-infrared2) bands. The MS image has a spatial resolution of 2.0 m and the PAN image has a spatial resolution of 0.5 m. Both of the two datasets include 400 × 400 pixels of MS image and 1600 × 1600 pixels of PAN image.

B. Experimental setup

In the synthetic experiments, the datasets of IKONOS and Worldview-2 were upscaled by a factor of four to generate the synthetic MS and PAN images, where the upscaling factor was equal to the spatial ratio between the MS and PAN images. Specifically, the spatial degradation of MS and PAN images was achieved using the same PSF based on the Gaussian filter in equation (15) with the spatial factor of four. For the IKONOS dataset, the spatial resolutions of the synthetic MS and PAN images is 16 m and 4 m, and the original 4 m MS images were used as reference images. For the two Worldview-2 datasets,

the synthetic MS and PAN images have a spatial resolution of 8 m and 2 m, where the original 2 m MS images were used as reference. Four quantitative indices were used for the validation of the resultant sharpened MS images against the reference MS images of the three datasets [11]: the relative global dimensional synthesis error (ERGAS), universal image quality index (UIQI), spectral angle mapper (SAM) and Q2n indexes (Q4 for four bands MS image and Q8 for eight bands MS image [77]).

In the real experiment, the spatial ratio was set to four in all the three datasets to produce the full resolution MS images (1 m for IKONOS MS image and 0.5 m for Worldview-2 MS image) for various pansharpening algorithms. Following this strategy, however, there are no reference FR MS images with which to validate the resultant sharpened MS images from the IKONOS and Worldview-2 datasets. In this situation, the *quality w/ no reference* (QNR) index [78] calculated by the integration of spatial (D_S) and spectral (D_λ) distortion indices were employed, and it was widely used to perform the quality evaluation at the original spatial resolution of the

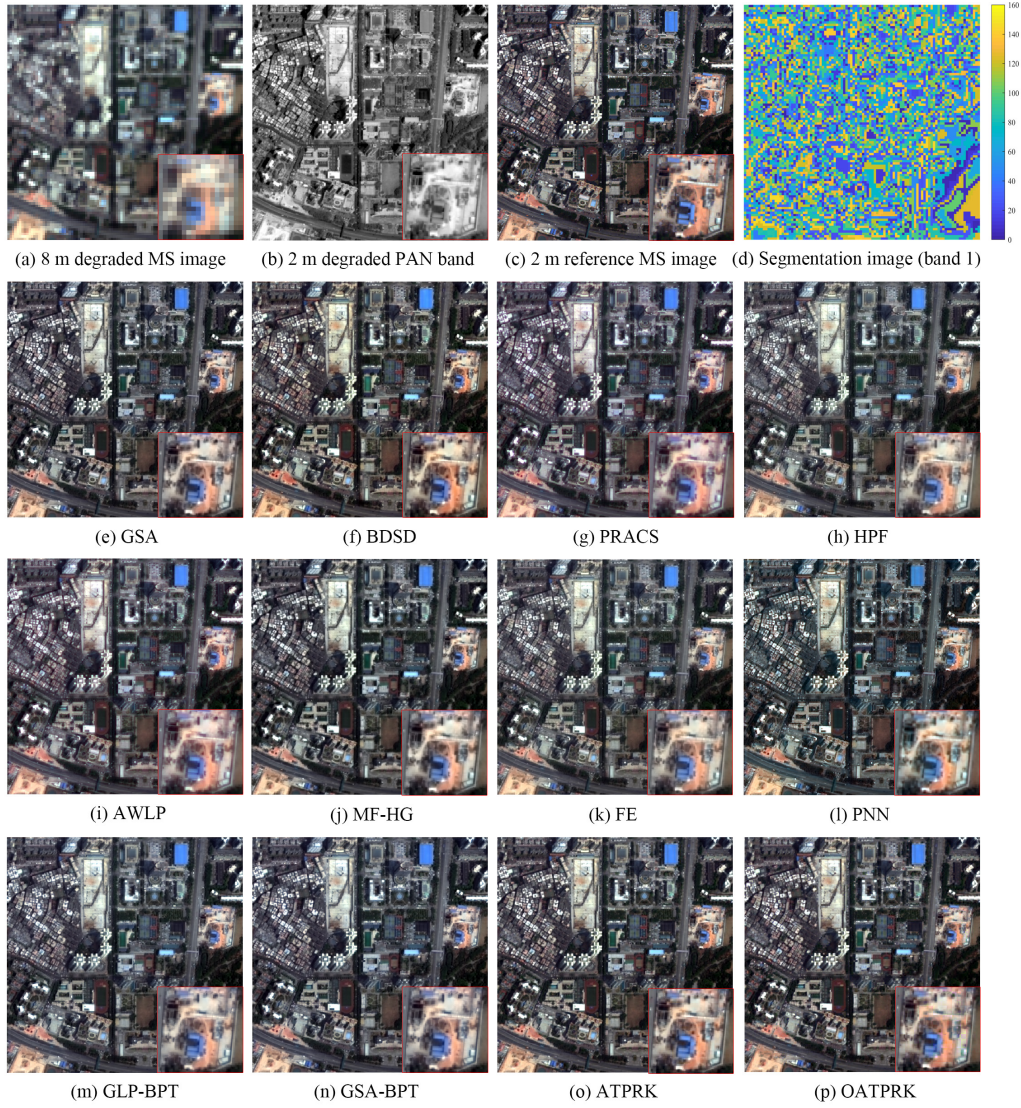|			
(a) 8 m degraded MS image	(b) 2 m degraded PAN band	(c) 2 m reference MS image	(d) Segmentation image (band 1)
(e) GSA	(f) BDSD	(g) PRACS	(h) HPF
(i) AWLP	(j) MF-HG	(k) FE	(l) PNN
(m) GLP-BPT	(n) GSA-BPT	(o) ATPRK	(p) OATPRK

Fig. 6. Results of different pansharpening algorithms for the Worldview-2 urban dataset (RGB: bands 532).

satellite sensor images. Smaller D_S and D_λ indicate that the fused FR MS has smaller spatial and spectral distortion. QNR is an integrated index, and the higher the QNR index, the better the quality of the sharpened FR MS images.

Nine benchmark pansharpening algorithms, including GSA [15], BDSD [65], partial replacement adaptive component substitution (PRACS) [16], HPF [18], AWLP [22], morphological operators based pansharpening (MF-HG) [37], optimal deconvolution filter estimation (FE) [36], PNN [26], GLP based on the binary partition tree (GLP-BPT) [62], GSA based on the binary partition tree (GSA-BPT) [62], ATPRK based on PSF [41, 79], were used as comparisons against the proposed OATPRK approach in both of the synthetic and real experiments. The implementation of the GSA, BDSD, PRACS, HPF and AWLP pansharpening algorithms was based on the MATLAB code provided in [78] , while the codes of MF-HG, FE, PNN, GLP-BPT and GSA-BPT were found in the platform of "open remote sensing" (https://openremotesensing.net/). The best results of ATPRK were obtained according to the parameter settings in [79].

C. Synthetic experiment

In this synthetic experiment, three datasets were used to evaluate the sharpened FR MS images at the reduced spatial resolution. Figs. 4-6 represent the fusion results of the various pansharpening algorithms for the three datasets, respectively. The corresponding quantitative assessment (based on four indices of UIQI, ERGAS, SAM, Q4/Q8) of the various pan-sharpening results is reported in Table II.

1) *IKONOS*: Fig. 4 represents the results of the various pansharpening algorithms illustrated by the IKONOS dataset, and Table II reports the corresponding quantitative assessment. The study area is covered mainly by urban, farmland, and forest. The synthetic two-band CR MS images for OATPRK were segmented into 145 objects (clusters) as shown in Fig. 4(d). PRACS produced a result with the lowest UIQI and highest ERGAS values and its color was not consistent to the reference image; moreover, the boundaries of buildings and farmlands were spatially blurred, and many spatial details were lost. Similar phenomenon can also be observed in the result of

TABLE III
QUANTITATIVE ASSESSMENT OF THE REAL PAN-SHARPENING RESULTS FOR THE IKONOS AND WORLDVIEW-2 DATASETS IN REAL EXPERIMENT.

	IKONOS			Worldview-2 rural			Worldview-2 urban		
	D_S	D_λ	QNR	D_S	D_λ	QNR	D_S	D_λ	QNR
Ideal	0	0	1	0	0	1	0	0	1
GSA	0.0325	0.0846	0.8856	0.0697	0.1475	0.7931	0.0607	0.0968	0.8483
BDSD	0.0340	0.0366	0.9307	0.0791	0.0861	0.8416	0.0229	0.0372	0.9407
PRACS	0.0310	0.0910	0.8808	**0.0255**	0.1037	0.8735	0.0275	0.0771	0.8975
HPF	0.0574	0.0766	0.8704	0.1047	0.1613	0.7509	0.0545	0.0714	0.8779
AWLP	0.0610	0.0841	0.8600	0.1198	0.1790	0.7227	0.0652	0.0808	0.8593
MF-HG	0.0800	0.1081	0.8206	0.1307	0.1820	0.7112	0.0863	0.1003	0.8221
FE	0.0592	0.0825	0.8632	0.1123	0.1690	0.7377	0.0574	0.0786	0.8685
PNN	0.1220	**0.0094**	0.8698	0.1108	0.1416	0.7633	0.0569	0.0406	0.9048
GLP-BPT	0.0368	0.0787	0.8874	0.1338	0.2159	0.6792	0.0598	0.0817	0.8634
GSA-BPT	0.0276	0.0833	0.8914	0.1086	0.1872	0.7246	0.0615	0.0969	0.8476
ATPRK	0.0404	0.0190	0.9414	0.0398	0.0824	0.8811	**0.0196**	0.0427	0.9386
OATPRK	**0.0265**	0.0166	**0.9573**	0.0776	**0.0293**	**0.8954**	0.0366	**0.0198**	**0.9443**

GSA. Although the boundaries of the results of HPF, AWLP, MF-HG and FE became a bit clearer, many spatial details were still seriously missing. For the result of BDSD, it has better accuracy values than most of the other methods, and the boundaries of various land covers became similar to the reference data, but the spectral distortion was serious. Spatial details in the result of PNN were well presented and the boundaries of many objects were enhanced, but serious spectral distorting was happened in the sharpened MS image (see Fig. 4(l)). GLP-BPT and GSA-BPT produced similar results that contained more spatial details than the CS- and MRA-based pansharpening algorithms, but the accuracy values for the result of GSA-BPT were a bit better than those of GLP-BPT. For the geostatistical based pansharpening algorithm of ATPRK, it produced result with color that consistent to the reference image, but the boundaries were also spatially blurred, like those of HPF and AWLP, MF-HG and FE. This is because this study site is covered by different land cover types (objects), and they have different spectral reflectance characteristics, which making it difficult for the linear regression in ATPRK to capture the spatial details of various objects. However, for the proposed OATPRK, as unique regression model was performed for each of the objects, it produced the most accurate result in terms of the quantitative comparison, and the spatial artifacts and blurred boundaries were reduced, and more spatial details were exploited, which is the most similar to the reference image. Compared with GSA, BDSD, PRACS, HPF, AWLP, MF-HG, FE, PNN, GLP-BPT, GSA-BPT and ATPRK, the UIQI value of OATPRK result increased by 10.52%, 2.61%, 14.10%, 7.49%, 6.08%, 4.44%, 5.77%, 5.69%, 5.39%, 4.78% and 6.55%, respectively, which demonstrates the superiority of OATPRK for fusing the IKONOS dataset.

2) *Worldview-2 rural:* Fig. 5 shows the sharpened MS images produced by various pansharpening algorithms for the Worldview-2 rural dataset, and Table II lists the corresponding quantitative assessment. This study site is located in a north rural area of Hong Kong and its land cover are dominated by water, trees, bare soil and roads. The synthetic two-band MS image was segmented into 145 objects as shown in Fig. 5(d). Among the CS-based pansharpening algorithms, BDSD produced results with greater spectral distortion as seen by

visual inspection of the water color of various ponds, and the result of BDSD presented more spatial details than that of GSA. With regard to the HPF and AWLP results, they also produced results with spectral distortion. PRACS produced the worst accurate result, as it had spatially blurred boundaries in the result, and it is hard to see the road line along with the ponds. For the variational optimization based pansharpening algorithms, MF-HG and FE produced similar results, while clear spatial details (e.g. the road line) were presented well in the PNN result, but its color was seriously distorted. Although clear spatial details were well presented in the results of GLP-BPT and GSA-BPT, the color of sharpened MS images was distorted, such as the color of vegetation was whitening. As shown in Fig. 5(j), ATPRK reduced the spectral distortion, but failed to restore spectral detail in many cases, and many spatial artifacts appeared around the buildings and road lines. Since OATPRK allows spatial variation in the regression relation between objects, spectral distortion was obviously reduced, as shown in Fig. 5(k). Meanwhile, spatial detail was restored in the OATPRK result. Compared with other methods, the accuracy of OATPRK increased for all indices. Compared with GSA, BDSD, PRACS, HPF, AWLP, MF-HG, FE, PNN, GLP-BPT, GSA-BPT and ATPRK, the UIQI value of OATPRK increased by 4.04%, 2.14%, 8.24%, 3.93%, 3.78%, 2.43%, 3.08%, 3.11%, 2.51%, 2.54% and 2.01%, respectively, which demonstrates the superiority of OATPRK for fusing the Worldview-2 dataset at rural area.

3) *Worldview-2 urban*: Fig. 6 shows the sharpened MS images produced by various pansharpening algorithms for the Worldview-2 urban dataset, and Table II lists the corresponding quantitative assessment. Considering that this study site is in the city center of Shenzhen in China and dominated by impervious surfaces and trees, the clustering number was set to 145 in OATPRK to achieve the best performance, and Fig. 6(d) displays the segmented image. Among the CS- and MRA-based pansharpening algorithms, GSA produced results with greater accuracy than the others. Similar to that of the Worldview-2 rural dataset, the variational optimization based pansharpening algorithms of MF-HG and FE produced similar results and spatial details were not that clear, obvious color (spectral) distortion was observed in the results of HPF and PNN, and PRACS produced the worst accurate result compared to the other methods, as the boundaries in the PRACS result were seriously spatially blurred. Besides GSA, the segmentation-based algorithms of GLP-BPT and GSA-BPT
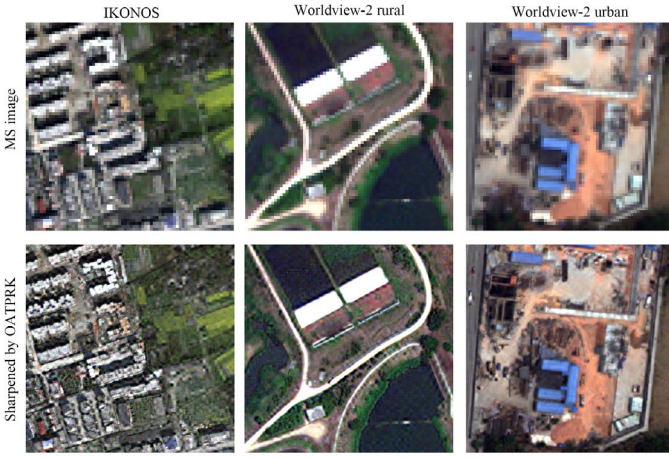
Fig. 7. Original subset MS images (4 m for IKONOS image, and 2 m for Worldview-2 images) and sharpened MS images (1 m for IKONOS image, and 0.5 m for Worldview-2 images) by using OATPRK for the IKONOS, Worldview-2 rural and urban datasets for the real experiment.

produced better results than the CS-, MRA- and variational optimization based pansharpening methods. With a similar trend observed by the IKONOS and Worldview-2 datasets, ATPRK produced a result with better accuracy values than all of the CS- and MRA-based pansharpening algorithms, but the OATPRK result achieved the greatest accuracy. Compared with GSA, BDSD, PRACS, HPF, AWLP, MF-HG, FE, PNN, GLP-BPT, GSA-BPT and ATPRK, the UIQI value of OATPRK increased by 0.50%, 1.02%, 2.85%, 2.59%, 1.93%, 1.21%, 1.93%, 1.19%, 0.77%, 0.70% and 0.25%, respectively, which demonstrates the superiority of OATPRK for fusing the Worldview-2 dataset at the urban area.

D. Real experiment

To test the performance of OATPRK in a real application, Table III reports the accuracy assessment of full resolution sharpened MS images for the IKONOS and Worldview-2 datasets. The spatial ratio of the three datasets are four, and then the sharpened MS images have a spatial size of 1600×1600 pixels. It is noted the results of BDSD and PRACS have larger QNR values than all of the other CS- and MRA-based pansharpening algorithms, but they cannot produce results with both high D_λ and D_s values. Specially, PRACS is able to retain more accurate spectral information and its results have smaller D_s values, while the results of BDSD have smaller D_λ value and more spatial information was exploited. For the variational optimization based pansharpening algorithm of MF-HG, its result has the lowest QNR values in all of the three datasets, and the highest D_s and D_λ values in the Worldview-2 rural and urban datasets. Compared with MF-HG, the performance of FE has an improvement. The QNR values for the result of PNN are higher than the other two variational optimization based algorithms; moreover, its result has the lowest D_λ values in the Worldview-2 rural dataset, which shows the advantage of PNN in exploiting spatial detail information. For the segmentation-based algorithms of GLP-BPT and GSA-BPT, they produced better results than most of the CS-, MRA- and variational optimization based
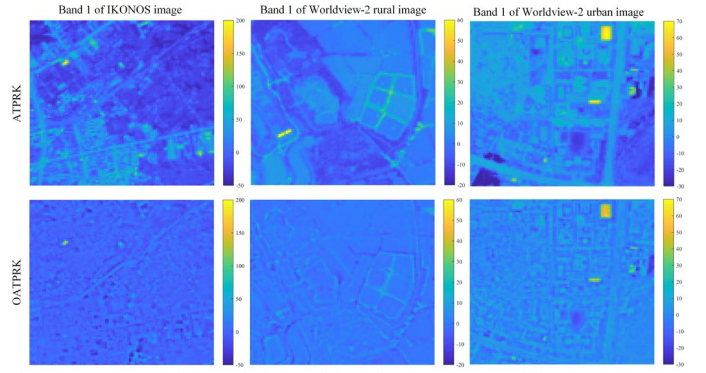

Fig. 8. Coarse spatial resolution residual images generated by the regression models in ATPRK and OATPRK for the band 1 of IKONOS, Worldview-2 rural and urban images.

pansharpening methods in the IKONOS dataset experiment; but their performance was not that good in the Worldview-2 rural and urban datasets experiments. Similar to the synthetic experiment, the geostatistical methods of ATPRK and the proposed OATPRK produce greater accuracies than most of the CS- and MRA-based methods, but the superiority of ATPRK was not consistently achieved. For example, in the Worldview-2 urban dataset experiment, the BDSD result had a larger QNR value than those of ATPRK. However, in all three experiments, the proposed OATPRK method produced results with the smallest D_λ and largest QNR values, which demonstrates the efficiency and superiority of OATPRK in a real application, and the subset MS images of the three datasets produced by OATPRK is shown in Fig. 7.

IV. DISCUSSION

A. Relationship between ATPRK and OATPRK

ATPRK holds the key advantage of perfectly preserving the spectral properties of the observed MS images, and is composed of the linear regression and residual downscaling parts. For the proposed OATPRK, it inherits the key advantage of ATPRK and an additional image segmentation part was developed, where the regression model was built with respect to the variation of different objects. It is noteworthy that if the clustering number is set to 1, then the proposed OATPRK approach becomes the basic ATPRK. That is, OATPRK should not have lower accuracy than ATPRK, and in general, as the above experiments demonstrated, where the clustering number is greater than one, OATPRK is expected to be superior. In general, the main difference between ATPRK and OATPRK is the regression part, and then the superiority of OATPRK also belongs to this part. Fig. 8 is used here to illustrate the coarse spatial resolution residual images generated by the regression models of ATPRK and OATPRK for the band 1 of IKONOS, Worldview-2 rural and urban images. It is obvious to see that the residual images generated by the regression model of ATPRK include high values for various land covers, especially for the impervious surface and water bodies. However, the values in the residual images of OATPRK are mostly close to 0,

TABLE IV
INFLUENCE OF THE SEGMENTATION METHODS USED IN OATPRK FOR THREE
DATASETS IN SYNTHETIC EXPERIMENT.

Segmentation method	IKONOS		Worldview-2 rural		Worldview-2 urban	
	UIQI	ERGAS	UIQI	ERGAS	UIQI	ERGAS
K-means	0.8284	2.6020	0.9179	2.0973	0.9367	1.9724
FCM	0.8410	2.7129	0.9227	2.0900	0.9370	1.9734
FCM_S	0.9269	1.7396	0.9616	1.4749	0.9713	1.3491

which means it has captured more spatial details about various land covers in the PAN bands in the regression model, and this is why it has better performance. Therefore, the superiority of OATPRK is proved to be mainly coming from the object-based regression part. If the extreme case is there are not too many objects in the images, such as the optimal clustering number is 1, OATPRK will at least have the same performance as ATPRK. But for most of the case, the optimal clustering number is more than one, and then OATPRK would have a better performance than ATPRK.

B. Influence of image segmentation methods

For the proposed OATPRK approach, the image segmentation method based on FCM with spatial constraints (FCM_S) was applied and achieved a good performance with respect to the synthetic and real experiments. Generally, the image segmentation method has a significant influence on the performance of OATPRK. Table IV is used to illustrate the influence of segmentation methods on OATPRK in terms of the synthetic experiment, in which the clustering number K was set to 6 for all of three datasets. In particular, another two benchmark image segmentation methods, including the K-means [80] and ordinary FCM [71], were used as comparisons. As listed in Table IV, in the three datasets, K-means based OATPRK produced results with smaller UIQI values than those of FCM-based OATPRK. Because the K-means is a crisp or hard segmentation method, it is, therefore, not suitable for the mixed pixel problem in satellite sensor images [81]. Compared with K-means, FCM holds the key advantage of retaining more information in satellite sensor images by introducing the concept of fuzzy membership. The FCM-based OATPRK produced results with an improvement in the UIQI values compared with K-means, but the ERGAS values for the results based on K-means and FCM are similar. As both the FCM and K-means do not consider the spatial information from image context and would, thus, be sensitive to noise in FR satellite sensor images [82]. By introducing the spatial constraints on ordinary FCM, FCM_S based OATPRK achieved an obvious increase in accuracy. This is because that the image segmentation results of FCM and K-means in OATPRK always have numerous isolated pixels, which will result in many abnormal sharpened pixels with extremely high or low values. However, FCM_S in OATPRK can produce spatially smooth segmentation image and the isolated pixels will be eliminated, and then the abnormal sharpened pixels will
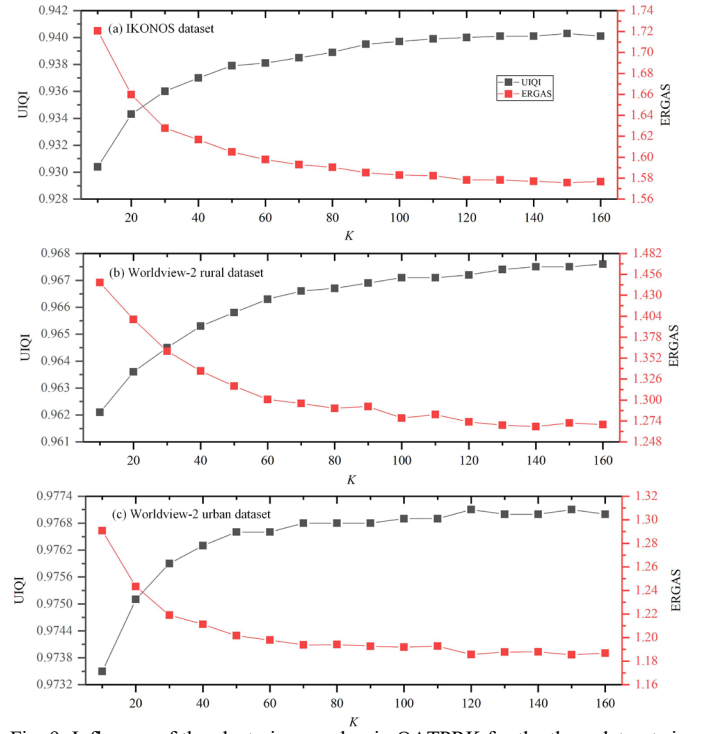

Fig. 9. Influence of the clustering number in OATPRK for the three datasets in the synthetic experiment.

not appear in the sharpened MS image. This indicates that for the proposed OATPRK, image segmentation method considering the spatial context information is essential for it to have a good performance. However, it is noteworthy that when the clustering number K was set to be a low value (e.g. 2), the performance of the three segmentation methods would be similar, because more spatial context information is considered when the number of clusters is very small. Besides the FCM_S, FCM and K-means, many other image segmentation algorithms, such as the BPT used in [62], simple linear iterative clustering (SLIC) method used in [63] and the Convolutional Neural Network (CNN) used in [83], can also be integrated with the OATPRK algorithm. In the future, any other efficient image segmentation algorithms considered the spatial context information are suggested to be applied in OATPRK to have a better performance.

C. Influence of clustering number

For OATPRK, the clustering number K in the image segmentation part (FCM_S) is an important parameter. Fig. 9 is used here to illustrate the influence of clustering number (K) on the performance of OATPRK for the three datasets in the synthetic experiment, in which K was in the range of 10-160 with an interval of 10. It is noteworthy that when K is equal to 1, OATPRK would be the same as ATPRK. For all of the three datasets, the accuracies of OATPRK results increased rapidly when K increased from 10 to 60, and when K was continuously increasing from 60 to 100, the increase rate of UIQI value and the decrease rate of ERGAS value become significantly slight; finally, there will be no obvious change of UIQI and ERGAS

TABLE V
INFLUENCE OF THE LOCAL WINDOW SIZE IN OATPRK FOR THE THREE DATASETS IN THE SYNTHETIC EXPERIMENT

w	IKONOS		Worldview-2 rural		Worldview-2 urban	
	UIQI	ERGAS	UIQI	ERGAS	UIQI	ERGAS
3	0.9412	1.5760	0.9680	1.2794	0.9761	1.1953
5	0.9401	1.5765	0.9675	1.2693	0.9770	1.1861
7	0.9387	1.5872	0.9656	1.2983	0.9766	1.2091
9	0.9376	1.5911	0.9640	1.3175	0.9761	1.2249
11	0.9363	1.6062	0.9630	1.3289	0.9757	1.2382
13	0.9358	1.6088	0.9620	1.3435	0.9756	1.2369
15	0.9347	1.6136	0.9611	1.3518	0.9757	1.2358
17	0.9343	1.6201	0.9605	1.3486	0.9759	1.2321
19	0.9336	1.6195	0.9597	1.3666	0.9759	1.2307
21	0.9326	1.6279	0.9596	1.3651	0.9757	1.2341

values when K is larger than 100. This is because if K is set too small, the spatial variation of the fused image would not be represented; if K is too large, there would be too many isolated pixels in the segmentation image, which could lead to unreliable regression models in OATPRK for the isolated pixels. This demonstrates that OATPRK is sensitive to the variation of clustering number K. For the IKONOS and Worldview-2 datasets, the best UIQI values were achieved when K was around 150. It is noted that the computation time of the OATPRK algorithm would increase with the increase of K, and there should be a trade-off between the accuracy and computation cost for applying the proposed OATPRK algorithm in the real situation.

D. Influence of local window size

The local window size w in equation (2) is an important parameter used to take into account the spatial contextual information in the segmentation-based regression model of OATPRK. To have a validation of the influence of local window size w on the performance of OATPRK, its value was set in a range of 3-21 with an interval of 2, and the resultant sharpened MS images' UIQI and ERGAS values are listed in Table V. In terms of the trend, the impact of local window size is not that obvious for the result of OATPRK. When w was 3 or 5, OATPRK produced results with the highest UIQI and lowest ERGAS values. However, with continuous increase of w, the results' UIQI value becomes higher, while the ERGAS value has a continuous decrease for all of the three datasets. This is because when the value of w is set too large, many small-scale objects will be over eliminated, and the corresponding spectral and spatial information will not be well captured, which indicates that a small window size (e.g. 3 or 5) is suggestted for the implement of OATPRK. Moreover, it is noted that a small window size will have a low computation cost.

E. Computational efficiency

For the geostatistical methods of ATPRK and the proposed OATPRK, Table VI reports the computational cost for the synthetic IKONOS dataset, in which the clustering number K was set to 145. All of the algorithms used in this

TABLE VI
COMPUTATION COST OF THE GEOSTATISTICAL PANSHARPENING METHODS FOR SYNTHETIC IKONOS DATASET.

	Part 1 (s)	Part 2 (s)	Part 3 (s)	Total (s)
	Segmentation	Regression	Downscaling residuals	
ATPRK	0	0.29	8.16	8.45
OATPRK(K=145)	63.34	8.40	8.16	79.90
OATPRK(K=80)	37.62	3.64	8.16	49.42

research were implemented on the MATLAB R2018a version on an Intel(R) Core (TM) i9-9880H Processer at 2.30 GHz. The residual downscaling part is the same for the three methods and took more than 8.16 s. The regression process in ATPRK took less than 0.29 s, but OATPRK took around 8.40 s. This is because it fits a unique regression model for each of the objects and the time consumed increases exponentially with increasing size of satellite sensor image. The unique part of image segmentation in OATPRK took around 63.34 s, and thus, the total time consumed with OATPRK is 79.90 s for the IKONOS dataset. However, the computation time would decrease obviously if the number of K is set to a small value. As listed in Table VI, the part of "image segmentation" in OATPRK takes most of the time. In fact, computational efficiency of the "image segmentation" part is mainly affected by the clustering number, and the running time would be greatly reduced with the decrease of clustering number (e.g. when K=80). Meanwhile, as shown in Fig. 9, the accuracies of the resultant sharpened MS images still stay at a high level when the clustering number is 80, and the increase trend of UIQI is not obvious with the continuous increase of clustering number.

V. CONCLUSION

Most pansharpening approaches extract spatial detail from PAN image at the per-pixel scale. However, with an increase in spatial resolution, the spatial complexity of objects may increase exponentially and, thus, earth surface objects in PAN image may not be represented well on a per-pixel basis. To provide a solution to this problem, this paper developed a simple yet efficient object-based geostatistical pansharpening method, termed OATPRK, which overcomes the traditional, per-pixel based limitations of geostatistical pansharpening method. The image segmentation and object-based regression stages of OATPRK can be integrated to extract accurate and full spatial information from the PAN band for each band of CR MS image, while the image segmentation section could be replaced by any other powerful image segmentation/clustering algorithms; on the other hand, the residual downscaling stage in OATPRK can make the result consistent to the original CR MS image and avoid spectral distortion. Synthetic and real experiments based on IKONOS and Worldview-2 datasets were used to validate the performance of OATPRK against other eleven state-of-the-art benchmark pansharpening algorithms. OATPRK achieved sharpened MS images with consistently the greatest accuracies by comparing with other

CS-, MRA-, variational optimization, and segmentation based pansharpening algorithms, where the results contained more accurate spatial details and had less spectral distortions. In future research, effort would be given to the automatic estimation of optimal parameters for the use in OATPRK. Moreover, the MATLAB code of the proposed OATPRK algorithm will be freely available in the ResearchGate platform (https://www.researchgate.net/profile/Yihang_Zhang2).

REFERENCES

[1] G. Simone, A. Farina, F. C. Morabito *et al.*, "Image fusion techniques for remote sensing applications," *Inf. Fusion,* vol. 3, no. 1, pp. 3-15, 2002/03/01/, 2002.

[2] F. Gao, J. Masek, M. Schwaller *et al.*, "On the blending of the Landsat and MODIS surface reflectance: predicting daily Landsat surface reflectance," *IEEE Trans. Geosci. Remote Sens.,* vol. 44, no. 8, pp. 2207-2218, 2006.

[3] Z. Zhu, and C. E. Woodcock, "Continuous change detection and classification of land cover using all available Landsat data," *Remote Sens. Environ.,* vol. 144, pp. 152-171, 2014/03/25/, 2014.

[4] H. Shen, X. Meng, and L. Zhang, "An Integrated Framework for the Spatio–Temporal–Spectral Fusion of Remote Sensing Images," *IEEE Trans. Geosci. Remote Sens.,* vol. 54, no. 12, pp. 7135-7148, 2016.

[5] X. Zhu, E. H. Helmer, F. Gao *et al.*, "A flexible spatiotemporal method for fusing satellite images with different resolutions," *Remote Sens. Environ.,* vol. 172, pp. 165-177, 2016/01/01/, 2016.

[6] X. Li, F. Ling, G. M. Foody *et al.*, "Generating a series of fine spatial and temporal resolution land cover maps by fusing coarse spatial resolution remotely sensed images and fine spatial resolution land cover maps," *Remote Sens. Environ.,* vol. 196, pp. 293-311, 2017/07/01/, 2017.

[7] Y. Zhang, G. Foody, F. Ling *et al.*, "Spatial-temporal fraction map fusion with multi-scale remotely sensed images," *Remote Sens. Environ.,* vol. 213, pp. 162-181, 08/01, 2018.

[8] E. Pardo-Igúzquiza, M. Chica-Olmo, and P. M. Atkinson, "Downscaling cokriging for image sharpening," *Remote Sens. Environ.,* vol. 102, no. 1, pp. 86-98, 2006/05/30/, 2006.

[9] C. Thomas, T. Ranchin, L. Wald *et al.*, "Synthesis of Multispectral Images to High Spatial Resolution: A Critical Review of Fusion Methods Based on Remote Sensing Physics," *IEEE Trans. Geosci. Remote Sens.,* vol. 46, no. 5, pp. 1301-1312, 2008.

[10] T. Ranchin, B. Aiazzi, L. Alparone *et al.*, "Image fusion—the ARSIS concept and some successful implementation schemes," *ISPRS-J. Photogramm. Remote Sens.,* vol. 58, no. 1, pp. 4-18, 2003/06/01/, 2003.

[11] G. Vivone, L. Alparone, J. Chanussot *et al.*, "A Critical Comparison Among Pansharpening Algorithms," *IEEE Trans. Geosci. Remote Sens.,* vol. 53, no. 5, pp. 2565-2586, May, 2015.

[12] H. Yésou, Y. Besnus, and J. Rolet, "Extraction of spectral information from Landsat TM data and merger with SPOT panchromatic imagery — a contribution to the study of geological structures," *ISPRS-J. Photogramm. Remote Sens.,* vol. 48, no. 5, pp. 23-36, 1993/10/01/, 1993.

[13] N. Koutsias, M. Karteris, and E. Chuvieco, "The Use of Intensity-Hue-Saturation Transformation of Landsat5 Thematic Mapper Data for Burned Land Mapping," *Photogramm Eng. Remote Sensing,* vol. 66, pp. 829-839, 07/01, 2000.

[14] A. Laben Craig, and V. Brower Bernard, *Process For Enhancing The Spatial Resolution Of Multispectral Imagery Using Pan-sharpening*, US US 6011875 A,to EASTMAN KODAK CO, 2000.

[15] B. Aiazzi, S. Baronti, and M. Selva, "Improving Component Substitution Pansharpening Through Multivariate Regression of MS + Pan Data," *IEEE Trans. Geosci. Remote Sens.,* vol. 45, no. 10, pp. 3230-3239, 2007.

[16] J. Choi, K. Yu, and Y. Kim, "A New Adaptive Component-Substitution-Based Satellite Image Fusion by Using Partial Replacement," *IEEE Trans. Geosci. Remote Sens.,* vol. 49, no. 1, pp. 295-309, 2011.

[17] X. X. Zhu, and R. Bamler, "A Sparse Image Fusion Algorithm With Application to Pan-Sharpening," *IEEE Trans. Geosci. Remote Sens.,* vol. 51, no. 5, pp. 2827-2836, 2013.

[18] J. P. Chavez, S. Sides, and J. Anderson, "Comparison of Three Different Methods to Merge Multiresolution and Multispectral Data: Landsat TM and SPOT Panchromatic," *Photogramm Eng. Remote Sensing,* vol. 57, pp. 265-303, 03/01, 1991.

[19] J. G. Liu, "Smoothing Filter-based Intensity Modulation: A spectral preserve image fusion technique for improving spatial details," *Int. J. Remote Sens.,* vol. 21, no. 18, pp. 3461-3472, 2000/01/01, 2000.

[20] B. Aiazzi, L. Alparone, S. Baronti *et al.*, "MTF-tailored Multiscale Fusion of High-resolution MS and Pan Imagery," *Photogramm Eng. Remote Sensing,* vol. 72, pp. 591-596, 05/01, 2006.

[21] M. J. Shensa, "The discrete wavelet transform: wedding the a trous and Mallat algorithms," *IEEE Transactions on Signal Processing,* vol. 40, no. 10, pp. 2464-2482, 1992.

[22] X. Otazu, M. Gonzalez-Audicana, O. Fors *et al.*, "Introduction of sensor spectral response into image fusion methods. Application to wavelet-based methods," *IEEE Trans. Geosci. Remote Sens.,* vol. 43, no. 10, pp. 2376-2385, 2005.

[23] K. Amolins, Y. Zhang, and P. Dare, "Wavelet based image fusion techniques — An introduction, review and comparison," *ISPRS-J. Photogramm. Remote Sens.,* vol. 62, no. 4, pp. 249-263, 2007/09/01/, 2007.

[24] S. Li, and B. Yang, "A New Pan-Sharpening Method Using a Compressed Sensing Technique," *IEEE Trans. Geosci. Remote Sens.,* vol. 49, no. 2, pp. 738-746, 2011.

[25] X. X. Zhu, D. Tuia, L. Mou *et al.*, "Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources," *IEEE Trans. Geosci. Remote Sens.,* vol. 5, no. 4, pp. 8-36, 2017.

[26] G. Masi, D. Cozzolino, L. Verdoliva *et al.*, "Pansharpening by Convolutional Neural Networks," *Remote Sens.,* vol. 8, no. 7, pp. 594, 2016.

[27] R. Fei, J. Zhang, J. Liu *et al.*, "Convolutional Sparse Representation of Injected Details for Pansharpening," *IEEE Geosci. Remote Sens. Lett.,* vol. 16, no. 10, pp. 1595-1599, 2019.

[28] J. Hu, P. Hu, X. Kang *et al.*, "Pan-Sharpening via Multiscale Dynamic Convolutional Neural Network," *IEEE Trans. Geosci. Remote Sens.,* pp. 1-14, 2020.

[29] Y. Wei, Q. Yuan, H. Shen *et al.*, "Boosting the Accuracy of Multispectral Image Pansharpening by Learning a Deep Residual Network," *IEEE Geosci. Remote Sens. Lett.,* vol. 14, no. 10, pp. 1795-1799, 2017.

[30] M. Jiang, H. Shen, J. Li *et al.*, "A differential information residual convolutional neural network for pansharpening," *ISPRS-J. Photogramm. Remote Sens.,* vol. 163, pp. 257-271, 2020/05/01/, 2020.

[31] Y. Zhang, C. Liu, M. Sun *et al.*, "Pan-Sharpening Using an Efficient Bidirectional Pyramid Network," *IEEE Trans. Geosci. Remote Sens.,* vol. 57, no. 8, pp. 5549-5563, 2019.

[32] Q. Xie, M. Zhou, Q. Zhao *et al.*, "MHF-Net: An Interpretable Deep Network for Multispectral and Hyperspectral Image Fusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 1-1, 2020.

[33] G. Vivone, "Robust Band-Dependent Spatial-Detail Approaches for Panchromatic Sharpening," *IEEE Trans. Geosci. Remote Sens.,* vol. 57, no. 9, pp. 6421-6433, 2019.

[34] F. Palsson, J. R. Sveinsson, and M. O. Ulfarsson, "A New Pansharpening Algorithm Based on Total Variation," *IEEE Geosci. Remote Sens. Lett.,* vol. 11, no. 1, pp. 318-322, 2014.

[35] L. Deng, G. Vivone, W. Guo *et al.*, "A Variational Pansharpening Approach Based on Reproducible Kernel Hilbert Space and Heaviside Function," *IEEE Transactions on Image Processing,* vol. 27, no. 9, pp. 4330-4344, 2018.

[36] G. Vivone, M. Simões, M. D. Mura *et al.*, "Pansharpening Based on Semiblind Deconvolution," *IEEE Trans. Geosci. Remote Sens.,* vol. 53, no. 4, pp. 1997-2010, 2015.

[37] R. Restaino, G. Vivone, M. D. Mura *et al.*, "Fusion of Multispectral and Panchromatic Images Based on Morphological Operators," *IEEE Transactions on Image Processing,* vol. 25, no. 6, pp. 2882-2895, 2016.

[38] E. Pardo-Iguzquiza, V. F. Rodríguez-Galiano, M. Chica-Olmo *et al.*, "Image fusion by spatially adaptive filtering using downscaling cokriging," *ISPRS-J. Photogramm. Remote Sens.,* vol. 66, no. 3, pp. 337-346, 2011/05/01/, 2011.

[39] P. M. Atkinson, E. Pardo-Iguzquiza, and M. Chica-Olmo, "Downscaling Cokriging for Super-Resolution Mapping of Continua in Remotely Sensed

Images," *IEEE Trans. Geosci. Remote Sens.,* vol. 46, no. 2, pp. 573-580, 2008.

[40] M. H. R. Sales, C. M. Souza, and P. C. Kyriakidis, "Fusion of MODIS Images Using Kriging With External Drift," *IEEE Trans. Geosci. Remote Sens.,* vol. 51, no. 4, pp. 2250-2259, 2013.

[41] Q. Wang, W. Shi, P. M. Atkinson *et al.*, "Downscaling MODIS images with area-to-point regression kriging," *Remote Sens. Environ.,* vol. 166, pp. 191-204, 2015/09/01/, 2015.

[42] Q. Wang, W. Shi, and P. M. Atkinson, "Area-to-point regression kriging for pan-sharpening," *ISPRS-J. Photogramm. Remote Sens.,* vol. 114, pp. 151-165, 2016/04/01/, 2016.

[43] Y. Zhang, P. M. Atkinson, F. Ling *et al.*, "Spectral–Spatial Adaptive Area-to-Point Regression Kriging for MODIS Image Downscaling," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.,* vol. 10, no. 5, pp. 1883-1896, 2017.

[44] C. Pohl, and J. L. Van Genderen, "Review article Multisensor image fusion in remote sensing: Concepts, methods and applications," *Int. J. Remote Sens.,* vol. 19, no. 5, pp. 823-854, 1998/01/01, 1998.

[45] V. Walter, "Object-based classification of remote sensing data for change detection," *ISPRS-J. Photogramm. Remote Sens.,* vol. 58, no. 3, pp. 225-238, 2004/01/01/, 2004.

[46] M. Wurm, H. Taubenböck, M. Schardt *et al.*, "Object-based image information fusion using multisensor earth observation data over urban areas," *Int. J. Image Data Fusion,* vol. 2, no. 2, pp. 121-147, 2011/06/01, 2011.

[47] G. J. Hay, K. O. Niemann, and G. F. McLean, "An object-specific image-texture analysis of H-resolution forest imagery," *Remote Sens. Environ.,* vol. 55, no. 2, pp. 108-122, 1996/02/01/, 1996.

[48] T. Blaschke, "Object based image analysis for remote sensing," *ISPRS-J. Photogramm. Remote Sens.,* vol. 65, no. 1, pp. 2-16, 2010/01/01/, 2010.

[49] T. Blaschke, G. J. Hay, M. Kelly *et al.*, "Geographic Object-Based Image Analysis – Towards a new paradigm," *ISPRS-J. Photogramm. Remote Sens.,* vol. 87, pp. 180-191, 2014/01/01/, 2014.

[50] J. J. Lewis, R. J. O'Callaghan, S. G. Nikolov *et al.*, "Pixel- and region-based image fusion with complex wavelets," *Inf. Fusion,* vol. 8, no. 2, pp. 119-130, 2007/04/01/, 2007.

[51] D. Rodríguez-Esparragón, J. Marcello, F. Eugenio *et al.*, "Object-based quality evaluation procedure for fused remote sensing imagery," *Neurocomputing,* vol. 255, pp. 40-51, 2017/09/13/, 2017.

[52] S. W. Myint, P. Gober, A. Brazel *et al.*, "Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery," *Remote Sens. Environ.,* vol. 115, no. 5, pp. 1145-1161, 2011/05/15/, 2011.

[53] H. Costa, G. M. Foody, and D. S. Boyd, "Supervised methods of image segmentation accuracy assessment in land cover mapping," *Remote Sens. Environ.,* vol. 205, pp. 338-351, 2018/02/01/, 2018.

[54] Y. Ge, Y. Chen, S. Li *et al.*, "Vectorial boundary-based sub-pixel mapping method for remote-sensing imagery," *Int. J. Remote Sens.,* vol. 35, no. 5, pp. 1714-1768, 2014/03/04, 2014.

[55] Y. Chen, Y. Ge, G. B. M. Heuvelink *et al.*, "Object-Based Superresolution Land-Cover Mapping From Remotely Sensed Imagery," *IEEE Trans. Geosci. Remote Sens.,* vol. 56, no. 1, pp. 328-340, 2018.

[56] G. M. Foody, "Crop classification from airborne synthetic aperture radar data," *Int. J. Remote Sens.,* vol. 9, no. 4, pp. 655-668, 1988/04/01, 1988.

[57] H. Costa, G. Foody, and D. Boyd, "Using mixed objects in the training of object-based image classifications," *Remote Sens. Environ.,* vol. 190, pp. 188-197, 03/01, 2017.

[58] Z. Chen, Z. Zhao, P. Gong *et al.*, "A new process for the segmentation of high resolution remote sensing imagery," *Int. J. Remote Sens.,* vol. 27, no. 22, pp. 4991-5001, Nov 20, 2006.

[59] B. Wuest, and Y. Zhang, "Region based segmentation of QuickBird multispectral imagery through band ratios and fuzzy comparison," *ISPRS-J. Photogramm. Remote Sens.,* vol. 64, no. 1, pp. 55-64, 2009/01/01/, 2009.

[60] U. C. Benz, P. Hofmann, G. Willhauck *et al.*, "Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information," *ISPRS-J. Photogramm. Remote Sens.,* vol. 58, no. 3, pp. 239-258, 2004/01/01/, 2004.

[61] D. Liu, and F. Xia, "Assessing object-based classification: advantages and limitations," *Remote Sens. Lett.,* vol. 1, no. 4, pp. 187-194, 2010/12/01, 2010.

[62] R. Restaino, M. D. Mura, G. Vivone *et al.*, "Context-Adaptive Pansharpening Based on Image Segmentation," *IEEE Trans. Geosci. Remote Sens.,* vol. 55, no. 2, pp. 753-766, 2017.

[63] H. Hallabia, H. Hamam, and A. B. Hamida, "An Optimal Use of SCE-UA Method Cooperated With Superpixel Segmentation for Pansharpening," *IEEE Geosci. Remote Sens. Lett.*, pp. 1-5, 2020.

[64] G. Vivone, S. Marano, and J. Chanussot, "Pansharpening: Context-Based Generalized Laplacian Pyramids by Robust Regression," *IEEE Trans. Geosci. Remote Sens.,* vol. 58, no. 9, pp. 6152-6167, 2020.

[65] A. Garzelli, F. Nencini, and L. Capobianco, "Optimal MMSE Pan Sharpening of Very High Resolution Multispectral Images," *IEEE Trans. Geosci. Remote Sens.,* vol. 46, no. 1, pp. 228-236, 2008.

[66] H. Li, L. Jing, Y. Tang *et al.*, "An Image Fusion Method Based on Image Segmentation for High-Resolution Remotely-Sensed Imagery," *Remote Sens.,* vol. 10, no. 5, pp. 790, 2018.

[67] C. Songcan, and Z. Daoqiang, "Robust image segmentation using FCM with spatial constraints based on new kernel-induced distance measure," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics),* vol. 34, no. 4, pp. 1907-1916, 2004.

[68] S. C. Chen, and D. Q. Zhang, "Robust image segmentation using FCM with spatial constraints based on new kernel-induced distance measure," *IEEE Trans. Syst. Man Cybern. Part B-Cybern.,* vol. 34, no. 4, pp. 1907-1916, 2004.

[69] K.-S. Chuang, H.-L. Tzeng, S. Chen *et al.*, "Fuzzy c-means clustering with spatial information for image segmentation," *Comput. Med. Imaging Graph.,* vol. 30, no. 1, pp. 9-15, 2006/01/01/, 2006.

[70] W. Cai, S. Chen, and D. Zhang, "Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation," *Pattern Recognit.,* vol. 40, no. 3, pp. 825-838, 2007/03/01/, 2007.

[71] J. K. Udupa, and S. Samarasekera, "Fuzzy Connectedness and Object Definition: Theory, Algorithms, and Applications in Image Segmentation," *Graph. Models Image Process.,* vol. 58, no. 3, pp. 246-261, 1996/05/01/, 1996.

[72] P. Addesso, G. Vivone, R. Restaino *et al.*, "A Data-Driven Model-Based Regression Applied to Panchromatic Sharpening," *IEEE Transactions on Image Processing,* vol. 29, pp. 7779-7794, 2020.

[73] P. M. Atkinson, "Downscaling in remote sensing," *Int. J. Appl. Earth Obs.,* vol. 22, no. Supplement C, pp. 106-114, 2013/06/01/, 2013.

[74] P. C. Kyriakidis, "A Geostatistical Framework for Area-to-Point Spatial Interpolation," *Geogr. Anal.,* vol. 36, no. 3, pp. 259-289, 2004.

[75] P. Goovaerts, "Kriging and Semivariogram Deconvolution in the Presence of Irregular Geographical Units," *Math Geosci.,* vol. 40, no. 1, pp. 101-128, January 01, 2008.

[76] P. C. Kyriakidis, and E.-H. Yoo, "Geostatistical Prediction and Simulation of Point Values from Areal Data," *Geogr. Anal.,* vol. 37, no. 2, pp. 124-151, 2005.

[77] A. Garzelli, and F. Nencini, "Hypercomplex Quality Assessment of Multi/Hyperspectral Images," *IEEE Geosci. Remote Sens. Lett.,* vol. 6, no. 4, pp. 662-665, 2009.

[78] L. Alparone, B. Alazzi, S. Baronti *et al.*, "Multispectral and panchromatic data fusion assessment without reference," *Photogramm. Eng. Remote Sens.,* vol. 74, no. 2, pp. 193-200, Feb, 2008.

[79] Q. Wang, W. Shi, Z. Li *et al.*, "Fusion of Sentinel-2 images," *Remote Sens. Environ.,* vol. 187, pp. 241-252, 2016/12/15/, 2016.

[80] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognit. Lett.,* vol. 31, no. 8, pp. 651-666, 2010/06/01/, 2010.

[81] N. Viovy, "Automatic Classification of Time Series (ACTS): A new clustering method for remote sensing time series," *Int. J. Remote Sens.,* vol. 21, no. 6-7, pp. 1537-1560, 2000/01/01, 2000.

[82] S. Bandyopadhyay, U. Maulik, and A. Mukhopadhyay, "Multiobjective Genetic Clustering for Pixel Classification in Remote Sensing Imagery," *IEEE Trans. Geosci. Remote Sens.,* vol. 45, no. 5, pp. 1506-1511, 2007.

[83] C. Du, and S. Gao, "Image Segmentation-Based Multi-Focus Image Fusion Through Multi-Scale Convolutional Neural Network," *IEEE Access,* vol. 5, pp. 15750-15761, 2017.

Yihang Zhang received his B.S. in Land Resource Management from the China University of Geosciences, Wuhan, China in 2012. He received his Ph.D. in Physical Geography from the Institute of Geodesy and Geophysics, Chinese Academy of Sciences, Wuhan, China in 2017.

He is currently an associate professor at the Institute of Geodesy and Geophysics, Innovation Academy for Precision Measurement Science and Technology, Chinese Academy of Sciences, Wuhan, China. During 2015-2016, he was a visiting Ph.D. student supervised by Peter M. Atkinson at Lancaster Environment Center, Faculty of Science and Technology, Lancaster University. His research interests are downscaling of remotely sensed imagery and forest cover mapping.

Peter M. Atkinson received the BSc degree in Geography from the University of Nottingham in 1986 and the PhD degree from the University of Sheffield (NERC CASE award with Rothamsted Experimental Station) in 1990. More recently, he received the MBA degree from the University of Southampton in 2012.

He is Dean of the Faculty of Science and Technology at Lancaster University. He was Professor of Geography at the University Southampton (for the last 21 years; 13 as Professor), where he is currently Visiting Professor. He is also Visiting Professor at Queen's University Belfast, UK. The main focus of his research is in remote sensing, geographical information science and spatial (and space-time) statistics applied to a range of environmental science and socio-economic problems. He has published around 200 peer-reviewed articles in international scientific journals and around 50 refereed book chapters. He has also edited nine journal special issues and eight books. He is Associate Editor for *Computers and Geosciences* and sits on the editorial boards of several further journals including *Geographical Analysis*, *Spatial Statistics*, the *International Journal of Applied Earth Observation and Geoinformation*, and *Environmental Informatics*. He sits on various international scientific committees.

Feng Ling received the B.S. degree in geology and the M.S. degree in geophysical prospecting and information technology from China University of Geosciences, Wuhan, China in 1999 and 2002, respectively, and the Ph.D. degree in system analysis and integration from Huazhong University of Science and Technology, Wuhan, China in 2006.

He is currently a professor with Institute of Geodesy and Geophysics, Innovation Academy for Precision Measurement Science and Technology, Chinese Academy of Sciences, Wuhan, China. His research interests include super-resolution mapping and spatial-temporal fusion of remotely sensed imagery. He has published more than 50 refereed journal articles. He received the 2016 Boeing Award for Best Scientific Paper in Image Analysis and Interpretation from the American Society for Photogrammetry and Remote Sensing (ASPRS) as the corresponding author. He has been supported by several talent programs, including the Hubei Province Natural Science Fund for Distinguished Young Scholars (2013) and the Distinguished Young Scientist Grant of the Chinese Academy of Sciences (2014).

Giles M. Foody (M'01–SM'10–F'13) earned the B.Sc. and Ph.D. degrees from the University of Sheffield, U.K., in 1983 and 1986, respectively.

He is currently Professor of Geographical Information Science and Head of the School of Geography at the University of Nottingham, U.K. His main research interests focus on the interface between remote sensing, ecology, and informatics. Topics of particular interest relate to image classification for land cover mapping and monitoring applications, addressing issues at scales ranging from the sub-pixel to global. His publication list includes 9 books and more than 195 refereed journal articles. Prof. Foody serves as founding editor-in-chief of *Remote Sensing Letters*. He holds additional editorial roles on eight other journals including *Landscape Ecology*, *International Journal of Remote Sensing*, *Remote Sensing of Environment*, *Geocarto International*, and the *International Journal of Applied Earth Observation and Geoinformation*. He has served on numerous committees and panels, notably the U.K.'s two most recent research assessment exercises: RAE2008 and REF2014. He was chair of the European Union COST Action on 'Mapping and the Citizen Sensor' (TD1202) which involves researchers from 30 countries.
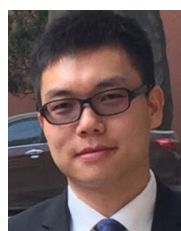
Qunming Wang (M'15) received the Ph.D. degree from The Hong Kong Polytechnic University, Hong Kong, in 2015. He was a Lecturer (Assistant Professor) with the Lancaster Environment Centre, Lancaster University, Lancaster, U.K., from 2017 to 2018. His three-year Ph.D. study was supported by the Hyper-Competitive Hong Kong Ph.D. Fellowship and his Ph.D. thesis was awarded as the outstanding thesis in the Faculty. He is currently a Professor with the College of Surveying and Geo-Informatics, Tongji University, Shanghai, China. He has authored over 40 peer-reviewed articles in international journals such as Remote Sensing of Environment, the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, and the ISPRS Journal of Photogrammetry and Remote Sensing. His current research interests include remote sensing, image processing, and geostatistics.

Dr. Wang serves as a reviewer for 30 international journals, including most of the international journals in remote sensing. He is an Associate Editor of Photogrammetric Engineering & Remote Sensing and Computers and Geosciences.

Yong Ge (M'14) received the Ph.D. degree in cartography and geographical information system from the Chinese Academy of Sciences (CAS), Beijing, China, in 2001.

She is a Professor with the State Key Laboratory of Resources and Environmental Information System, Institute of Geographical Sciences and Natural Resources Research, CAS. Her research activity focuses on spatial data analysis and data quality assessment. She has directed research in more than ten national projects. She is the author or coauthor of over 110 scientific papers published in refereed journals, one book, and six chapters in books; she holds six granted patents in improving the accuracy of information extraction from remotely sensed imagery. Dr. Ge has been involved in the organization of several international conferences and workshops. She is a steering committee member of International Spatial Accuracy Research Association(ISARA), and an editorial board member of Spatial Statistics (Elsevier).
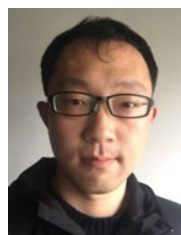
Xiaodong Li received his B.S. in Geographic Information System from the China University of Geosciences, Wuhan, China in 2006. He received his M.S. and Ph.D. in Physical Geography from the Institute of Geodesy and Geophysics, Chinese Academy of Sciences, Wuhan, China in 2009 and 2012, respectively.

He is currently an associate professor at the Institute of Geodesy and Geophysics, Innovation Academy for Precision Measurement Science and Technology, Chinese Academy of Sciences, Wuhan, China. He has been supported by the Youth Innovation Promotion Association CAS and the Hubei Province Natural Science Fund for Distinguished Young Scholars. His research interest is super-resolution mapping of remotely sensed imagery.

Yun Du received his B.S. in Geomorphology and Quaternary Geology from Nanjing University, Nanjing, China in 1989, and his M.S. in Physical Geography from the Institute of Geodesy and Geophysics, Chinese Academy of Sciences, Wuhan, China in 1992. He received his Ph.D. in Historical Geography from Wuhan University, Wuhan, China in 1999.

He is currently a professor at the Institute of Geodesy and Geophysics, Innovation Academy for Precision Measurement Science and Technology, Chinese Academy of Sciences, Wuhan, China. His research interests are remote sensing application and hydrology.