Information Loss-Guided Multi-Resolution Image Fusion

Qunming Wang, Wenzhong Shi, and Peter M. Atkinson

***Abstract*—Spatial downscaling is an ill-posed, inverse problem, and information loss (IL) inevitably exists in the predictions produced by any downscaling technique. The recently popularized area-to-point kriging (ATPK)-based downscaling approach can account for the size of support and the point spread function (PSF) of the sensor and, moreover, it has the appealing advantage of the perfect coherence property. In this paper, based on the advantages of ATPK and the conceptualization of IL, an IL-guided image fusion (ILGIF) approach is proposed. ILGIF uses the fine spatial resolution images acquired in other wavelengths to predict the IL in ATPK predictions based on the geographically weighted regression (GWR) model, which accounts for the spatial variation in land cover. ILGIF inherits all the advantages of ATPK and its prediction has perfect coherence with the original coarse spatial resolution data which can be demonstrated mathematically. ILGIF was validated using two datasets and was shown in each case to predict downscaled images more accurately than the compared benchmark methods.**

***Index Terms*—Image fusion, downscaling, geostatistics, information loss (IL), geographically weighted regression (GWR).**

I. INTRODUCTION

Downscaling is a process to increase the spatial resolutions of observations [1]. For remote sensing images, such a process involves the change-of-support problem (COSP), where the support is a geostatistical term meaning the space on which an observation or measurement is defined. The geostatistics-based area-to-point kriging (ATPK) technique is an effective solution to the COSP, which can predict a support that is smaller than that of the original data [2], [3]. ATPK was originally developed for census data (e.g., disease or health data) involving irregular geographical units (e.g., county) with different sizes and shapes [4]. Recently, the technique was popularized and extended to the remote sensing case which involves regular supports (pixels with the same size and shape) [5]. ATPK accounts for the size of support, spatial correlation, and the point spread function (PSF) of the sensor and has the appealing characteristic of perfect coherence with the original coarse spatial resolution data and, thus, it is an accurate method for downscaling [2].

Q. Wang is with the College of Surveying and Geo-Informatics, Tongji University, 1239 Siping Road, Shanghai 200092, China (e-mail: wqm11111@126.com).

W. Shi is with the Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Kowloon, Hong Kong.

P.M. Atkinson is with the Faculty of Science and Technology, Lancaster University, Lancaster LA1 4YR, UK; School of Geography, Archaeology and Palaeoecology, Queen's University Belfast, BT7 1NN, Northern Ireland, UK; and also with Geography and Environment, University of Southampton, Highfield, Southampton SO17 1BJ, UK (e-mail: pma@lancaster.ac.uk).

A. Information loss (IL) in downscaling

Downscaling is essentially an ill-posed, inverse problem, in which multiple plausible solutions can lead to an equally coherent recreation of the original coarse image. As a result, some of the required fine spatial resolution information cannot be recovered in the process, particularly for heterogeneous landscapes and boundaries between land cover types. That is, there is unavoidably information loss (IL) in downscaling solutions, where the terminology IL is defined as the gap between the ideal fine spatial resolution image (i.e., reference image) and actual downscaling solutions (e.g., those based on spatial prediction, e.g., using ATPK), as shown in Fig. 1. IL is defined in contrast to information gain (IG) which refers to the gain of the downscaling solution over the original coarse image. The relation between the input coarse image and the ideal downscaling solution can be summarized in Eq. (1). Although the objective of downscaling is to minimize the IL, such loss always exists and is never zero. If the IL can be predicted, it can compensate the ATPK-based predictions to achieve more accurate downscaling predictions.
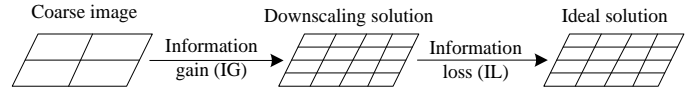

Fig. 1. The definition of IG and IL in image downscaling.

$$\text{Ideal solution} = \underbrace{\text{Coarse image} + \text{IG}}_{\text{Downscaling solution}} + \text{IL}. \qquad (1)$$

B. Potential solutions to IL prediction

1) Learning-based solution. For downscaling in real applications, the reference (i.e., the ideal solution) is always unavailable (otherwise there is no need for downscaling). Thus, the IL for the study area at the required fine spatial resolution cannot be predicted straightforwardly. A plausible solution to predict IL for a downscaling prediction is to find the relation between the downscaling prediction (or original coarse image, as input) and the IL (as output) based on training data, and apply the fitting model to the downscaling prediction of the study area. The training images need to be at the same spatial resolution as the target fine spatial resolution for downscaling, and more importantly, need to have a similar spatial pattern as the study area [6]. In most cases, there may not be easy access to such demanding training data. Alternatively, the fitting model could be predicted based on a self-example scheme [7]: the coarse image of the study area is upscaled to a coarser spatial resolution, and the original coarse image is treated as the ideal solution to calculate the IL. In this scheme, however, the IL is predicted at the original coarse spatial resolution. For remote sensing data, the spatial content can be different when the spatial resolution varies. For example, the roads and buildings are visible in a 5 m

spatial resolution image, but may 'disappear' at a coarser (e.g., 20 m) spatial resolution.

2) Multi-resolution image fusion-based solution. With the development of satellite sensors such as WorldView, QuickBird, IKONOS, SPOT, Landsat ETM+ and more recently, Sentinel-2 Multispectral Imager (MSI) [8], the Earth's surface can be observed at different spatial resolutions in different wavebands. The finer spatial resolution images in some wavebands (e.g., 15 m panchromatic (PAN) band in Landsat ETM+ or 10 m bands in Sentinel-2 MSI) have been used to guide the downscaling process for coarser spatial resolution images in other wavebands (e.g., 30 m multispectral bands in Landsat ETM+ or 20 m bands in Sentinel-2 MSI). This process is commonly known as multi-resolution image fusion in remote sensing, which has received increasing attention in recent years especially in relation to reliable monitoring.

C. Brief review of multi-resolution image fusion

Over the past decades, various multi-resolution image fusion methods have been developed. Two popular families are component substitution (CS) and multi-resolution analysis (MRA) [9]. The CS approach includes principal component analysis [10], band-dependent spatial-detail (BDSD) algorithm [11], Gram-Schmidt (GS) transformation [12], adaptive GS (GSA) [13], context-adaptive GSA (GSA-CA) [14] and partial replacement adaptive component substitution (PRACS) [15]. Common MRA examples are high-pass filtering [16], smoothing filter-based intensity modulation [17], *a trous* wavelet transform (ATWT) [18], *Additive Wavelet Luminance Proportional* (AWLP) [19], the generalized Laplacian pyramid with modulation transfer function-matched filter (MTF-GLP) [20], MTF-GLP with context-based decision (MTF-GLP-CBD) [21], and MTF-GLP with multiplicative injection model (MTF-GLP-HPM) [22]. Recently, sparse representation [23] and deep learning [24] based methods have also been developed for multi-resolution image fusion. It is beyond the scope of this paper to review the existing image fusion approaches explicitly, and several useful review articles exist [9], [25].

Multi-resolution image fusion methods were originally developed for the case of fusing a single PAN band (also termed pan-sharpening in remote sensing). Recently, Selva *et al.* [26] investigated the extension of the methods to the more general case of fusing more than one fine spatial resolution band, which is also termed "hypersharpening". Specifically, two schemes (i.e., the selected band and synthesized band schemes) are summarized for using multiple fine spatial resolution bands.

D. The proposed IL-guided image fusion (ILGIF) approach

With the availability of fine spatial resolution data in some wavebands, the IL in downscaling for these bands can be quantified by downscaling the coarse data (simulated by upscaling the known fine resolution data) and comparing the predictions with the known fine spatial resolution image. The fine spatial resolution bands can be treated as training data, and the IL in these bands can be used to predict the IL in downscaling coarse images in other wavebands. On this basis, a new IL-guided image fusion (ILGIF) approach is proposed for fusing multi-resolution images. Based on the ATPK solution to the COSP, the ILGIF prediction is the combination of the ATPK

prediction for the coarse image and the corresponding prediction for the IL.

According to one of the protocols in Wald *et al.* [27], any fused synthetic image, once degraded to its original spatial resolution, should be as identical as possible to the original coarse image. This has been a great challenge for the existing image fusion methods. As a new multi-resolution image fusion method based on a new conceptualization, ILGIF has the appealing merits of preserving perfectly the spectral property of the original coarse images (can be demonstrated mathematically), and thus, satisfies the aforementioned protocol. Moreover, ILGIF accounts for the PSF of the sensor and is easy to implement. ILGIF is suitable for fusion of PAN and multispectral images (i.e., the standard pan-sharpening problem) and fusion of multispectral and multi/hyperspectral images (i.e., where two groups of images are in different wavelength ranges).

The remainder of this paper is organized into four sections. Section II introduces briefly the principles of ILGIF. Section III provides the experimental results of two groups of datasets for validation of ILGIF. Further issues related to ILGIF and opening future research are discussed in Section IV. Finally, Section V concludes the paper.

II. Methods

A. Problem formulation

Let $Z_C^l(\mathbf{x}_i)$ be the measurements of pixel C centered at $\mathbf{x}_i$ ($i=1,…, M$, where M is the number of pixels) in coarse band l ($l=1,…, L$, where L is the number of coarse bands), and $Z_F^k(\mathbf{x}_j)$ be the measurements of pixel F centered at $\mathbf{x}_j$ ($j=1,…, MG^2$, where G is the spatial resolution (zoom) ratio) in fine band k ($k=1,…, K$, where K is the number of fine bands). Note that F and C represent the fine and coarse pixels, respectively. The objective of downscaling is to predict variables $Z_F^l(\mathbf{x})$ for all fine pixels in all L coarse bands. In the proposed ILGIF method, the process consists of ATPK-based downscaling and IL estimation. Denote the predictions of ATPK and IL as $\hat{Z}_{FA}^l(\mathbf{x})$ and $\hat{Z}_{FI}^l(\mathbf{x})$, the ILGIF prediction is

$$\hat{Z}_F^l(\mathbf{x}) = \hat{Z}_{FA}^l(\mathbf{x}) + \hat{Z}_{FI}^l(\mathbf{x}). \qquad (2)$$

The calculation of ATPK and IL predictions is detailed in the following Section II-B, and Sections II-C and -D.

Fig. 2 is the flowchart of the proposed ILGIF method, where a coarse band l is used as an example for illustration. The implementation of ILGIF is summarized by the following steps.

Step 1 For a coarse band l, it is downscaled to the fine spatial resolution using ATPK. This step is detailed in Section II-B.

Step 2 The ILs for the K fine bands in other wavebands are calculated, see Eqs. (8), (9) and (11). This step is detailed in Section II-C.

Step 3 The K weights transforming the ILs for the K fine bands to that for the coarse band are calculated using geographically weighted regression (GWR). This step is detailed in Section II-D.

Step 4 The IL for the coarse band is calculated (see Eq. (13)) and added to the ATPK prediction in step 1) (see Eq. (2)).

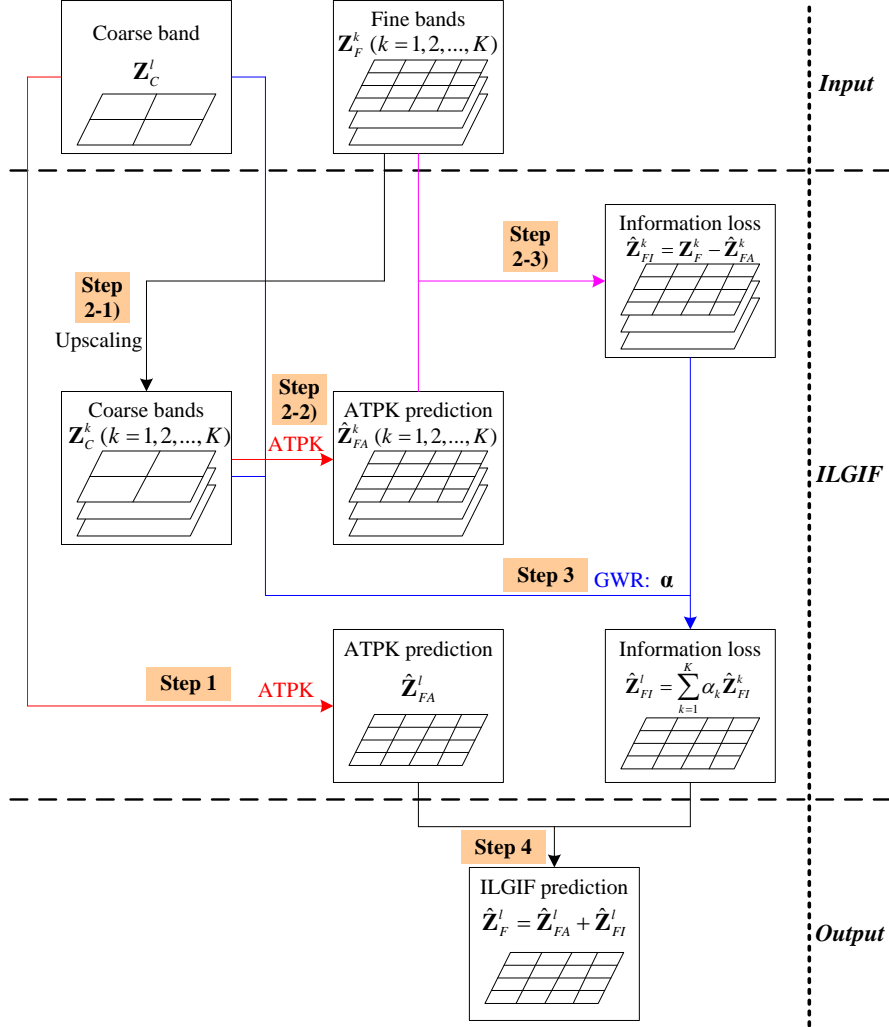Step 5 Steps 1-4 are performed for all *L* coarse bands.

Fig. 2. Flowchart of the proposed ILGIF method, where red and blue lines represent the ATPK and IL prediction processes, respectively. A coarse band *l* is used as the example and the process is implemented for each coarse band in turn.

B. Area-to-point kriging (ATPK)

For a fine pixel centered at $\mathbf{x}_0$ in band l, the ATPK-based downscaling prediction can be simply described as a linear combination of the neighboring coarse pixels

$$\hat{Z}_{FA}^l(\mathbf{x}_0) = \sum_{i=1}^{N} \lambda_i Z_C^l(\mathbf{x}_i), \text{ s.t. } \sum_{i=1}^{N} \lambda_i = 1 \quad (3)$$

where λ_i is the weight for the ith coarse neighboring pixel centered at $\mathbf{x}_i$ and N is the number of coarse neighbors. The N weights are calculated according to the kriging matrix below

$$\begin{bmatrix} \gamma_{CC}^l(\mathbf{x}_1,\mathbf{x}_1) & \cdots & \gamma_{CC}^l(\mathbf{x}_1,\mathbf{x}_N) & 1 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \gamma_{CC}^l(\mathbf{x}_N,\mathbf{x}_1) & \cdots & \gamma_{CC}^l(\mathbf{x}_N,\mathbf{x}_N) & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \cdot \\ \cdot \\ \cdot \\ \lambda_N \\ \theta \end{bmatrix} = \begin{bmatrix} \gamma_{FC}^l(\mathbf{x}_0,\mathbf{x}_1) \\ \cdot \\ \cdot \\ \cdot \\ \gamma_{FC}^l(\mathbf{x}_0,\mathbf{x}_N) \\ 1 \end{bmatrix}. \quad (4)$$

In (4), $\gamma_{CC}^l(\mathbf{x}_i,\mathbf{x}_j)$ is the coarse-to-coarse semivariogram between coarse pixels centered at $\mathbf{x}_i$ and $\mathbf{x}_j$ in band l,

$\gamma_{FC}^l(\mathbf{x}_0,\mathbf{x}_j)$ is the fine-to-coarse semivariogram between fine (to be predicted) and coarse pixels centered at $\mathbf{x}_0$ and $\mathbf{x}_j$ in band l, and θ is the Lagrange multiplier. Let $\mathbf{s}$ be the Euclidean distance between the centroids of any two pixels, $\gamma_{FF}^l(\mathbf{s})$ be the fine-to-fine semivariogram between two fine pixels, and $h_C^l(\mathbf{s})$ be the PSF for band l. $\gamma_{CC}^l(\mathbf{s})$ and $\gamma_{FC}^l(\mathbf{s})$ in (4) are calculated by convoluting $\gamma_{FF}^l(\mathbf{s})$ with the PSF $h_C^l(\mathbf{s})$ as follows

$$\gamma_{FC}^l(\mathbf{s}) = \gamma_{FF}^l(\mathbf{s}) * h_C^l(\mathbf{s}) \quad (5)$$

$$\gamma_{CC}^l(\mathbf{s}) = \gamma_{FF}^l(\mathbf{s}) * h_C^l(\mathbf{s}) * h_C^l(-\mathbf{s}) \quad (6)$$

where * is the convolution operator. The key issue becomes the estimation of the fine-to-fine semivariogram $\gamma_{FF}^l(\mathbf{s})$. If any prior spatial structure information at target fine spatial resolution for the band is available, it can be used readily for estimation. However, such information is not always available in reality. In this case, its estimation is achieved based on deconvolution, where the original coarse data are treated as areal data. The optimal solution to the fine-to-fine semivariogram is identified as the one that, when convolved with the PSF according to (6), is

the same as the areal semivariogram. Details of the several approaches for deconvolution can be found in the literature [4], [6].

An appealing advantage of ATPK is that the prediction has perfect coherence with the input coarse image. That is, once the ATPK prediction is upscaled to the original coarse resolution, it is exactly the same as the original coarse data [2], [3]

$$\hat{Z}_{FA}^l(\mathbf{x}) * h_C^l(\mathbf{x}) = Z_C^l(\mathbf{x}) . \tag{7}$$

C. Information Loss (IL)

As downscaling is an ill-posed, inverse problem, there exists unavoidable IL in ATPK prediction when compared with the ideal prediction (reference). The IL may not be an important problem for homogeneous landscapes, but it is crucial for restoration of heterogeneous landscapes with great spatial variation. For more reliable downscaling, it is important to predict the IL in ATPK predictions. In this paper, it is achieved using the K available fine spatial resolution images $\mathbf{Z}_F^1$, $\mathbf{Z}_F^2$,…, $\mathbf{Z}_F^K$ in other wavebands. The process is detailed below.

1) Upscaling the K fine bands to simulated coarse bands. Each fine spatial resolution image is upscaled to match the spatial resolution of the coarse image $\mathbf{Z}_C^l$

$$Z_C^k(\mathbf{x}) = Z_F^k(\mathbf{x}) * h_C^l(\mathbf{x}) . \tag{8}$$

2) ATPK-based downscaling for the simulated K coarse bands. ATPK is performed on the simulated coarse image $\mathbf{Z}_C^k$ for band k to downscale it back to the fine spatial resolution. Similarly to (3), the prediction for a fine pixel centered at $\mathbf{x}_0$ in band k is

$$\hat{Z}_{FA}^k(\mathbf{x}_0) = \sum_{i=1}^N \beta_i Z_C^k(\mathbf{x}_i), \text{ s.t. } \sum_{i=1}^N \beta_i = 1 \tag{9}$$

in which β_i is the weight for the ith coarse neighbor. The weights calculated in the same way as in (5). Based on the perfect coherence property of ATPK, the simulated coarse image $\mathbf{Z}_C^k$ can be reproduced exactly when the ATPK prediction $\hat{Z}_{FA}^k(\mathbf{x})$ is upscaled to the coarse spatial resolution

$$\hat{Z}_{FA}^k(\mathbf{x}) * h_C^l(\mathbf{x}) = Z_C^k(\mathbf{x}) . \tag{10}$$

3) Calculating the ILs for the K fine bands. Since the reference for $\hat{Z}_{FA}^k(\mathbf{x}_0)$ is known, the IL in the ATPK prediction for the fine pixel at $\mathbf{x}_0$ in band k is quantified as

$$\hat{Z}_{FI}^k(\mathbf{x}_0) = Z_F^k(\mathbf{x}_0) - \hat{Z}_{FA}^k(\mathbf{x}_0) . \tag{11}$$

From (8) and (10), we can conclude an important property of the quantified IL: once it is upscaled to the coarse spatial resolution, it is zero

$$\begin{aligned} &\hat{Z}_{FI}^k(\mathbf{x}) * h_C^l(\mathbf{x}) \\ &= [Z_F^k(\mathbf{x}) - \hat{Z}_{FA}^k(\mathbf{x})] * h_C^l(\mathbf{x}) \\ &= Z_F^k(\mathbf{x}) * h_C^l(\mathbf{x}) - \hat{Z}_{FA}^k(\mathbf{x}) * h_C^l(\mathbf{x}) . \\ &= Z_C^k(\mathbf{x}) - Z_C^k(\mathbf{x}) \\ &= 0 \end{aligned} \tag{12}$$

4) Calculating the ILs for the L coarse bands. The ILs of the K fine bands are used to predict the ILs in the L coarse bands. Specifically, the IL in ATPK prediction for a pixel centered at $\mathbf{x}$ in coarse band l (i.e., $\hat{Z}_{FI}^l(\mathbf{x})$ in (2)), is assumed to be a linear combination of all K ILs in the K available fine bands

$$\hat{Z}_{FI}^l(\mathbf{x}) = \sum_{k=1}^K \alpha_k(\mathbf{x}) \hat{Z}_{FI}^k(\mathbf{x}) \tag{13}$$

where $\alpha_k(\mathbf{x})$ is the weight for the kth fine band. The weights are determined according to the relation between the coarse band l and K fine bands. That is, a larger weight will be assigned to band k if the relation between images $\mathbf{Z}_C^l$ and $\mathbf{Z}_C^k$ is larger, and *vice versa*.

As acknowledged widely, the spatial structure of land cover always varies spatially [28], [29]. For images composed of pixels, the relation between the coarse and fine bands is not fixed, but a function of the pixel. This requires a non-stationary spatially adaptive model to characterize the relation (e.g., a fitting model in a local window). Moreover, in the local window, pixels can exert different effects on the center, as their spatial distances to the center are not the same. Thus, it would be more reasonable to quantify their influence according to spatial distance. On this basis, the GWR model [30] is proposed to predict the weights in (13).

D. Geographically weighted regression (GWR)-based weight estimation

GWR has been used widely in spatial analysis [31] and data assimilation [32], [33]. The model can relate data from different sources or platforms. For example, GWR was used to relate field data (e.g., PM2.5) to satellite sensor data [32] and normalized difference vegetation index (NDVI) to rainfall [33]. GWR can also relate data acquired from the same platform, such as filling the missing data (due to could or scan line corrector (SLC)-off) in remote sensing images [34] using temporally close, complete data.

In this paper, GWR is applied for estimation of $\alpha_k(\mathbf{x})$ in (13) by relating remote sensing data acquired in different wavebands. GWR is a local model that accounts explicitly for the spatial non-stationarity between the dependent and independent variables. Moreover, it allows the contributions from neighbors to vary according to their distances to the center pixel [33]. With the coarse band l and K fine bands, the GWR model is constructed as

$$Z_C^l(\mathbf{x}) = \alpha_0(\mathbf{x}) + \sum_{k=1}^K \alpha_k(\mathbf{x}) Z_C^k(\mathbf{x}) . \tag{14}$$

In (14), $\alpha_0(\mathbf{x})$ is the intercept. Let $\mathbf{P}(\mathbf{x})$ be an $N_0 \times (K+1)$ matrix composed of the coarse pixel values of all K bands $\mathbf{Z}_C^1$, $\mathbf{Z}_C^2$,…, $\mathbf{Z}_C^K$ (produced according to (8)) in the local window (including N_0 pixels for each band) centered at $\mathbf{x}$, with the last column being a $N \times 1$ vector of ones; $\mathbf{Q}(\mathbf{x})$ be an $N_0 \times 1$ vector composed of the coarse pixel values of the local window centered at $\mathbf{x}$ in coarse band l; $\mathbf{W}(\mathbf{x})$ be an $N_0 \times N_0$ spatial weighting diagonal matrix. The K weights for the pixel, included in a $(K+1) \times 1$ vector, are predicted by

$$\boldsymbol{\alpha}(\mathbf{x}) = [\mathbf{P}(\mathbf{x})^\mathrm{T} \mathbf{W}(\mathbf{x}) \mathbf{P}(\mathbf{x})]^{-1} \mathbf{P}(\mathbf{x})^\mathrm{T} \mathbf{W}(\mathbf{x}) \mathbf{Q}(\mathbf{x}) . \tag{15}$$

As seen from (15), the matrices of $\mathbf{P}(\mathbf{x})$ and $\mathbf{Q}(\mathbf{x})$ constructed from a local window result in weights varying on a pixel basis, which can cope with spatial non-stationarity. Furthermore, the diagonal elements in $\mathbf{W}(\mathbf{x})$ ensure that pixels near to the location $\mathbf{x}$ have more influence on the prediction than

the further pixels [33]. They can be determined based on a bi-square function

$$w_i = \begin{cases} [1-(d_i/H)^2]^2, & \text{if } d_i < H \\ 0, & \text{otherwise} \end{cases} \quad (16)$$

in which d_i is the distance between the ith neighboring pixel and the center pixel at $\mathbf{x}$, and H is the bandwidth for the kernel.

E. Coherence property of ILGIF

As mentioned in (2), the final ILGIF prediction $\hat{Z}_F^l(\mathbf{x})$ is a combination of the ATPK prediction $\hat{Z}_{FA}^l(\mathbf{x})$ in (3) and IL prediction $\hat{Z}_{FI}^l(\mathbf{x})$ in (13). Combining (7), (13) and (12), we can derive the following important property of the ILGIF prediction

$$
\begin{aligned}
&\hat{Z}_F^l(\mathbf{x})*h_C^l(\mathbf{x}) \\
&= [\hat{Z}_{FA}^l(\mathbf{x}) + \hat{Z}_{FI}^l(\mathbf{x})]*h_C^l(\mathbf{x}) \\
&= \hat{Z}_{FA}^l(\mathbf{x})*h_C^l(\mathbf{x}) + \hat{Z}_{FI}^l(\mathbf{x})*h_C^l(\mathbf{x}) \\
&= Z_C^l(\mathbf{x}) + \sum_{k=1}^{K} \alpha_k(\mathbf{x})\hat{Z}_{FI}^k(\mathbf{x})*h_C^l(\mathbf{x}) . \\
&= Z_C^l(\mathbf{x}) + \sum_{k=1}^{K} \alpha_k(\mathbf{x})\times 0 \\
&= Z_C^l(\mathbf{x})
\end{aligned}
\quad (17)
$$

It means that once the ILGIF prediction is upscaled to the coarse spatial resolution, it is exactly the same as the original coarse input $\mathbf{Z}_C^l$, that is, it has the perfect coherence property. It should be noted that such a property is not affected by the specific value of weights $\alpha_k(\mathbf{x})$ and the specific form of PSF (as long as a consistent PSF is used in the whole process of ILGIF).

III. Experiments

A. Data and experimental setup

Two datasets were used for experimental validation of the proposed ILGIF method, including a WorldView-2 dataset and a Sentinel-2 dataset. The WorldView-2 dataset contains eight multispectral bands with a spatial resolution of 2 m and a PAN band with a spatial resolution of 0.5 m. The spatial sizes of the multispectral and PAN images are 400 by 400 pixels and 1600 by 1600 pixels, respectively. The data were acquired in April, 2011 and cover an urban area in Shenzhen, China.

The used Sentinel-2 dataset contains four 10 m bands and six 20 m bands. It was acquired on 18 August 2015. The study area is located in Verona, Italy, and is covered mainly by a mix of vegetation and urban fabric. The dataset has a spatial extent of 8 km by 8 km (400 by 400 pixels for 20 m bands and 800 by 800 pixels for 10 m bands).

For objective evaluation where fine spatial resolution data are required for examination, synthetic datasets were used (i.e., the reduced resolution case as termed in [9]). Specifically, for the WorldView-2 dataset, the eight 2 m multispectral bands and 0.5 m PAN band were upscaled to 8 m and 2 m by convolving them with a PSF, as shown in Fig. 3(a) and (b). Similarly, for the Sentinel-2 dataset, the six 20 m and four band 10 m bands were upscaled to 40 m and 20 m (see Fig. 5(a) and (b)). In all experiments, a Gaussian PSF was used and the standard deviation (size of the PSF width) was set to half of the coarse pixel size. The task of downscaling is to restore the eight 2 m WorldView-2 multispectral bands and six 20 m Sentinel-2 bands,

by fusing them with the synthesized 2 m WorldView-2 PAN band and four 20 m Sentinel-2 bands, respectively. The predictions were compared to the original 2 m WorldView-2 bands and 20 m Sentinel-2 bands for objective evaluation. This scheme has been used commonly to evaluate downscaling approaches [35]. For clarity, we termed the experiments for the two datasets as pan-sharpening and multispectral sharpening.

Four CS methods (i.e., PRACS [15], GSA [13], GSA-CA [14] and BDSD [11]) and six MRA methods (i.e., ATWT [18], AWLP [19], MTF-GLP [20], MTF-GLP-CBD [21], MTF-GLP-HPM [22] and the recently developed morphological half gradient (MF-HG) [36]) were considered as benchmark methods. The CS and MRA approaches use a single fine band (e.g., PAN band) for the coarse bands. Thus, a single band needs to be extracted from the set of fine bands to adapt them for multispectral sharpening. Two schemes summarized in [26] (i.e., the selected band and synthesized band schemes) were considered in the experiments. With respect to the selected band scheme, for each coarse band, the fine band with the greatest correlation (quantified by CC) with it was selected. Regarding the synthesized band scheme, for each coarse band, a single fine band was synthesized as a linear combination of the available fine bands. The weights were determined using the multiple regression model built between the coarse band and all fine bands.

For quantitative evaluation, we used the correlation coefficient (CC), universal image quality index (UIQI), $Q2^n$ index [37], relative global-dimensional synthesis error (ERGAS) and spectral angle mapper (SAM). CC and UIQI were first calculated for each band, and the values for all bands were finally averaged. For $Q2^n$ and SAM, they were calculated for each pixel first and then averaged. Moreover, to measure the ability to honour the original coarse data, coherence (quantified by the CC) was used. More precisely, the fused image was upscaled to the original coarse spatial resolution and evaluated with the original coarse image based on CC.

B. Experiment on pan-sharpening

ATPK-based downscaling for the input 8 m coarse image (Fig. 3(a)) is an important first step of ILGIF. To illustrate the advantage of ATPK-based downscaling, it was compared to the classical polynomial interpolation (with 23 coefficients). Fig. 3(c) shows the polynomial interpolation result for Fig. 3(a). Compared with the ATPK result in Fig. 3(d), the polynomial interpolation result is more blurred and the gaps between the buildings cannot be restored satisfactorily. Table 1 lists the accuracies for the two methods, where the advantage of ATPK is obvious from the quantitative comparison. More precisely, ATPK increases the $Q2^n$ and UIQI by around 0.10 and 0.07, respectively. The more satisfactory performance of ATPK mainly lies in the ability to account for the size of support and PSF and more importantly, the preservation of the original data (i.e., coherence property, see Eq. (10)).

Table 1 Comparison between polynomial interpolation and ATPK for the WorldView-2 dataset (the bold values mean the most accurate results in each term)

	CC	UIQI	$Q2^n$	ERGAS	SAM	Coherence
Ideal	1	1	1	0	0	1
Polynomial	0.8197	0.7741	0.6885	3.1083	0.0516	0.9802
ATPK	**0.8552**	**0.8440**	**0.7923**	**2.7643**	**0.0453**	**0.9985**

ATPK was performed on the 8 m upscaled PAN image, and the 2 m IL (in units of digital number (DN)) produced by comparing to the 2 m reference PAN in Fig. 3(b) is shown in Fig. 3(e). It is seen that for the boundaries of the small-sized buildings, there exists relatively large uncertainty in downscaling. Based on GWR, the IL in Fig. 3(e) was then used for estimation of the IL in ATPK-based restoration of the 2 m multispectral bands (i.e., the prediction in Fig. 3(d)). By adding the IL to the ATPK prediction, the final ILGIF prediction was produced, as shown in Fig. 3(f). It is clear that by adding the IL, the smoothing effect in ATPK result was obviously reduced and much more spatial detail was reproduced. As a result, the ILGIF result is much more similar to the reference in Fig. 3(g).

The ten benchmark methods were implemented. For a clearer comparison with the results, all fused images were compared with the reference in Fig. 3(g) and produced the error maps in Fig. 4. It is clear visually that the proposed ILGIF method has the smallest error among all methods, especially for restoration of building boundaries (heterogeneous features).

Table 2 lists the quantitative assessment results for all 11 methods. Comparing to the results in Table 1, it is seen that the accuracies of the image fusion methods are greater than that for the method using only the input coarse image. For example, both CC and UIQI are increased by about 0.12 from ATPK to ILGIF. Focusing on the result in Table 2, GSA and GSA-CA have very similar performance and the ERGASs are smaller than 1.6. Both are more accurate than the other two CS methods (i.e., PRACS and BDSD). Among the MRA methods, MTF-GLP, MTF-GLP-CBD and MTF-GLP-HPM tend to be more accurate. However, the accuracies of both the CS and MRA methods are smaller than the proposed ILGIF method. ILGIF produced the largest CC, UIQI and $Q2^n$ and smallest ERGAS and SAM. Regarding coherence, ILGIF produced a value very close to the ideal value of 1, suggesting its perfect coherence property.

Table 2 Quantitative assessment for different methods for the WorldView-2 dataset (the bold values mean the most accurate results in each term)

	CC	UIQI	$Q2^n$	ERGAS	SAM	Coherence
Ideal	1	1	1	0	0	1
PRACS	0.9499	0.9268	0.9067	2.0261	0.0457	0.9940
GSA	0.9663	0.9585	0.9413	1.5557	0.0401	0.9958
GSA-CA	0.9662	0.9581	0.9409	1.5618	0.0402	0.9957
BDSD	0.9516	0.9503	0.9317	1.7869	0.0466	0.9767
ATWT	0.9573	0.9459	0.9196	1.7010	0.0392	0.9929
AWLP	0.9529	0.9398	0.9187	1.8211	0.0419	0.9923
MTF-GLP	0.9603	0.9502	0.9262	1.6410	0.0387	0.9930
MTF-GLP-CBD	0.9558	0.9554	0.9274	1.6625	0.0371	0.9928
MTF-GLP-HPM	0.9606	0.9506	0.9269	1.6447	0.0393	0.9929
MF-HG	0.9589	0.9530	0.9283	1.6277	0.0386	0.9930
ILGIF	**0.9709**	**0.9692**	**0.9489**	**1.3581**	**0.0337**	**0.9997**

Table 3 Quantitative assessment for different methods for the WorldView-2 dataset at full resolution (the bold values mean the most accurate results in each term)

	QNR	Coherence
Ideal	1	1
PRACS	**0.9781**	0.9909
GSA	0.9632	0.9914
GSA-CA	0.9529	0.9891
BDSD	0.9393	0.9615
ATWT	0.9660	0.9934
AWLP	0.9613	0.9931
MTF-GLP	0.9641	0.9935
MTF-GLP-CBD	0.9665	0.9943
MTF-GLP-HPM	0.9646	0.9935
MF-HG	0.9550	0.9945
ILGIF	0.9749	**0.9997**

All 11 methods were also implemented for the full resolution case, that is, fusion of the eight 2 m multispectral bands and the 0.5 m PAN band to create an eight-band, 0.5 m multispectral image. The Quality with No Reference (QNR) index [38] was used to evaluate the methods quantitatively. As claimed by Palsson *et al.* [38], consistency can also give reliable assessment of the relative performance of image fusion methods at full resolution and it tends to be superior to the commonly used QNR metrics. Thus, the coherence was also used here. The results for the two indices are shown in Table 3. Comparing the QNR values, the ILGIF can produce greater accuracy than the benchmark methods except PRACS. Checking the coherence values, however, ILGIF has the largest value, suggesting the result is the most accurate.

C. Experiment on multispectral sharpening

The 20 m downscaling results of polynomial interpolation and ATPK for the input 40 m coarse Sentinel-2 images in Fig. 5(a) are shown in Fig. 5(c) and (d), respectively. Again, ATPK can reproduce more spatial details. For example, in Fig. 5(d), the linear features of the urban fabric can be observed more clearly. The advantage is also supported by the quantitative assessment in Table 3. Furthermore, by adding ILs derived from the four 20 m bands to the APTK result, the produced ILGIF result Fig. 5(f) is more accurate and much closer to the reference in Fig. 5(c).

The error maps for all 11 methods are shown in Fig. 6. For each benchmark method, the results for both selected and synthesized band schemes are exhibited. As seen from the results, ILGIF has the smallest error among all cases, which can be observed clearly by checking the locations of the rivers. Table 4 also indicates that ILGIF produces greater accuracies than the ten benchmark methods, no matter whether the selected or synthesized band scheme is applied. More precisely, the CCs and UIQIs of the ten methods are below 0.99, but ILGIF produced a CC and UIQI of 0.99. The ERGASs of the ten methods are all above 2.5 (even exceeds 3.4 for PRACS with both schemes), but for ILGIF it is about 2. In addition, the coherence value of ILGIF is almost the ideal value of 1.

Table 4 Comparison between polynomial interpolation and ATPK for the Sentinel-2 dataset (the bold values mean the most accurate results in each term)

	CC	UIQI	$Q2^n$	ERGAS	SAM	Coherence
Ideal	1	1	1	0	0	1
Polynomial	0.9287	0.9135	0.8713	5.3082	0.0450	0.9909
ATPK	**0.9501**	**0.9489**	**0.9257**	**4.2934**	**0.0381**	**0.9991**

D. Analysis of alternatives for ATPK and GWR in ILGIF

To analyze the advantages of using ATPK and GWR in the proposed ILGIF method, different combinations of interpolation and IL estimation were performed for the two datasets. Table 6 shows the accuracies of four combinations:

1) Polynomial + GLR (global linear regression) (i.e., MTF-GLP with synthesized band scheme);

2) Polynomial + GWR;

3) ATPK + GLR;

4) ATPK + GWR (i.e., the proposed ILGIF method).

By comparing ATPK + GLR (or GWR) to Polynomial + GLR (or GWR), it is seen clearly that the accuracies of the two

ATPK-based methods are greater than the two polynomial-based methods for both datasets. For example, focusing on the results for the WorldView-2 dataset, the $Q2^n$ of ATPK + GWR is 0.0118 larger than those of Polynomial + GWR, while the $Q2^n$ of ATPK + GLR is 0.0260 larger than those of Polynomial + GLR. This means that the use of ATPK is more advantageous than polynomial interpolation in the image fusion problem, which is also consistent with the findings in Tables 1 and 4. When comparing ATPK (or Polynomial) + GWR to ATPK (or Polynomial) + GLR, it is observed that the two GWR-based methods are more accurate than the two GLR-based methods, suggesting the benefits of using the GWR scheme in IL estimation. Overall, ATPK + GWR produces the most accurate results among all four combinations.

Table 5 Quantitative assessment for different methods for the Sentinel-2 dataset (the bold values mean the most accurate results in each term)

		CC	UIQI	$Q2^n$	ERGAS	SAM	Coherence
	Ideal	1	1	1	0	0	1
Selected band	PRACS	0.9759	0.9690	0.9602	3.4023	0.0346	0.9951
	GSA	0.9787	0.9768	0.9705	3.0390	0.0334	0.9952
	GSA-CA	0.9783	0.9766	0.9702	3.0513	0.0351	0.9940
	BDSD	0.9722	0.9701	0.9585	3.3719	0.0340	0.9944
	ATWT	0.9767	0.9765	0.9696	3.1310	0.0375	0.9952
	AWLP	0.9759	0.9745	0.9674	3.2080	0.0413	0.9946
	MTF-GLP	0.9781	0.9758	0.9682	3.0607	0.0336	0.9961
	MTF-GLP-CBD	0.9766	0.9763	0.9699	3.2438	0.0387	0.9958
	MTF-GLP-HPM	0.9800	0.9776	0.9691	2.9222	0.0324	0.9965
	MF-HG	0.9790	0.9782	0.9704	2.9427	0.0341	0.9959
Synthesized band	PRACS	0.9694	0.9647	0.9599	3.7088	0.0415	0.9838
	GSA	0.9675	0.9669	0.9617	3.8208	0.0442	0.9794
	GSA-CA	0.9733	0.9731	0.9667	3.4471	0.0411	0.9847
	BDSD	0.9716	0.9698	0.9586	3.4237	0.0354	0.9933
	ATWT	0.9812	0.9810	0.9751	2.8379	0.0324	0.9962
	AWLP	0.9798	0.9785	0.9722	2.9764	0.0367	0.9955
	MTF-GLP	0.9823	0.9807	0.9749	2.7582	0.0304	0.9969
	MTF-GLP-CBD	0.9839	0.9836	0.9780	2.6502	0.0312	0.9974
	MTF-GLP-HPM	0.9817	0.9788	0.9712	2.8587	0.0317	0.9967
	MF-HG	0.9820	0.9817	0.9760	2.7240	0.0311	0.9964
ILGIF		**0.9901**	**0.9899**	**0.9861**	**2.0578**	**0.0263**	**0.9999**

Table 6 Comparison between different combinations of interpolation and IL estimation methods (the bold values mean the most accurate results in each term)

	WorldView-2						Sentinel-2					
	CC	UIQI	$Q2^n$	ERGAS	SAM	Coherence	CC	UIQI	$Q2^n$	ERGAS	SAM	Coherence
Ideal	1	1	1	0	0	1	1	1	1	0	0	1
Polynomial + GLR	0.9603	0.9502	0.9262	1.6410	0.0387	0.9930	0.9823	0.9807	0.9749	2.7582	0.0304	0.9969
Polynomial + GWR	0.9659	0.9622	0.9371	1.4919	0.0372	0.9966	0.9884	0.9882	0.9838	2.2552	0.0282	0.9994
ATPK+GLR	0.9707	0.9679	**0.9522**	1.3988	0.0355	**0.9997**	0.9891	0.9888	0.9849	2.1308	**0.0260**	**0.9999**
ATPK+GWR	**0.9709**	**0.9692**	0.9489	**1.3581**	**0.0337**	**0.9997**	**0.9901**	**0.9899**	**0.9861**	**2.0578**	0.0263	**0.9999**

(a)

(b)

(c)

(d)
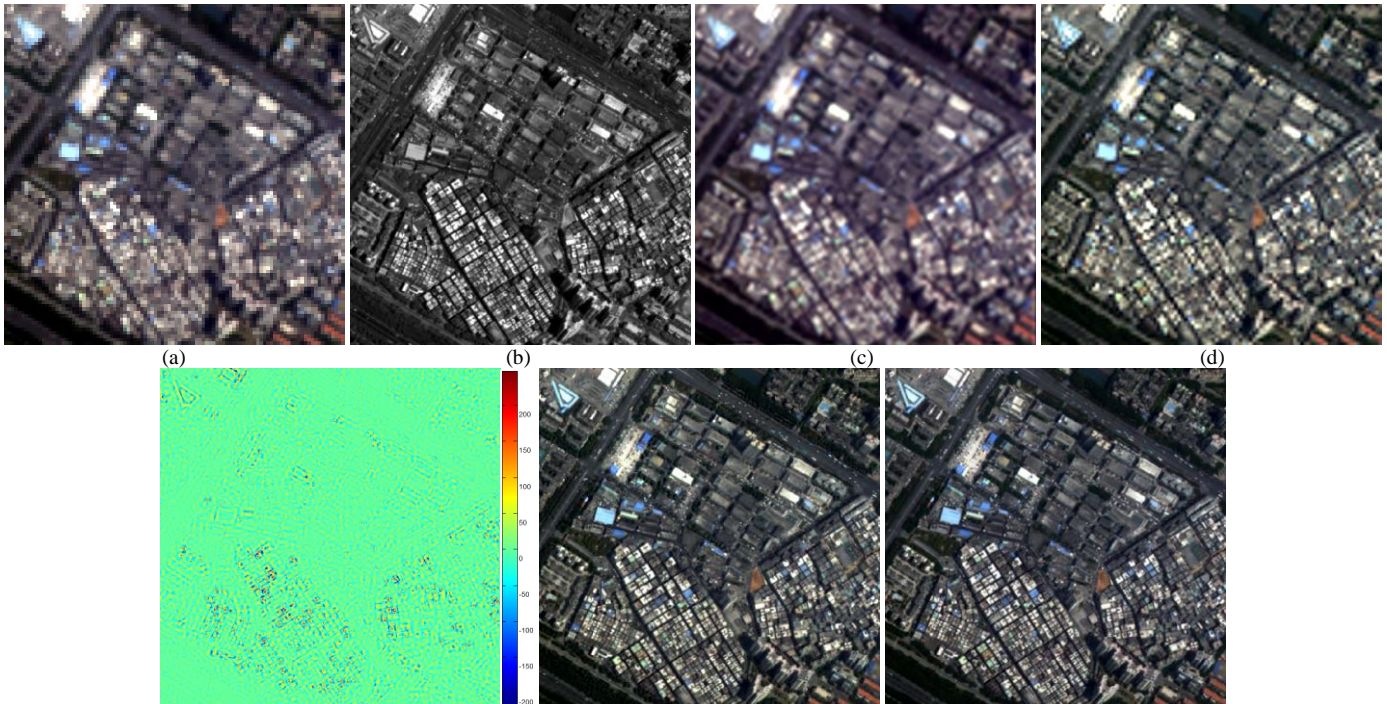
(e) (f) (g)

Fig. 3 Results for the WorldView-2 dataset (bands 4, 3 and 2 as RGB). (a) 8 m coarse multispectral image. (b) 2 m fine PAN image. (c) 2 m polynomial interpolation result. (d) 2 m ATPK result. (e) IL in downscaling 8 m PAN to 2 m. (f) 2 m ILGIF result. (g) 2 m reference.

(a) (b) (c) (d)

(e) (f) (g) (h)
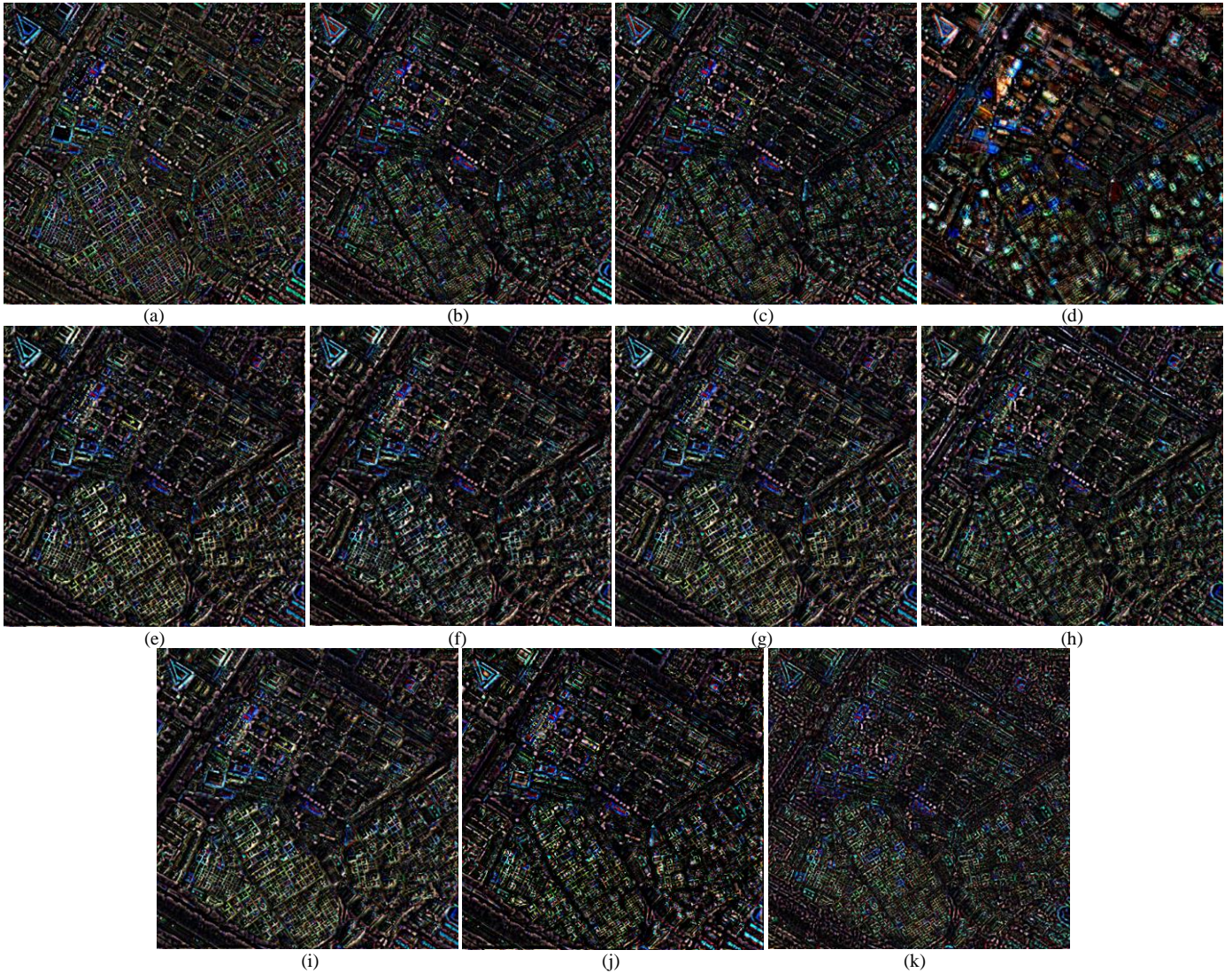
(i) (j) (k)

Fig. 4 Error maps for the different methods for the WorldView-2 dataset (bands 4, 3 and 2 as RGB). (a) PRACS. (b) GSA. (c) GSA-CA. (d) BDSD. (e) ATWT. (f) AWLP. (g) MTF-GLP. (h) MTF-GLP-CBD. (i) MTF-GLP-HPM. (j) MF-HG. (k) ILGIF.
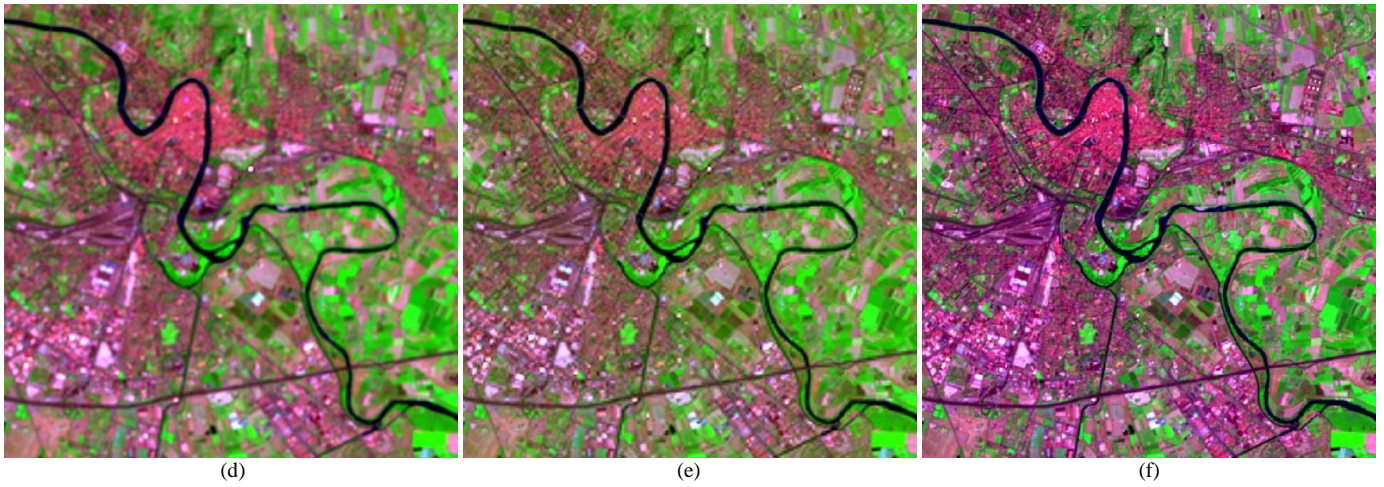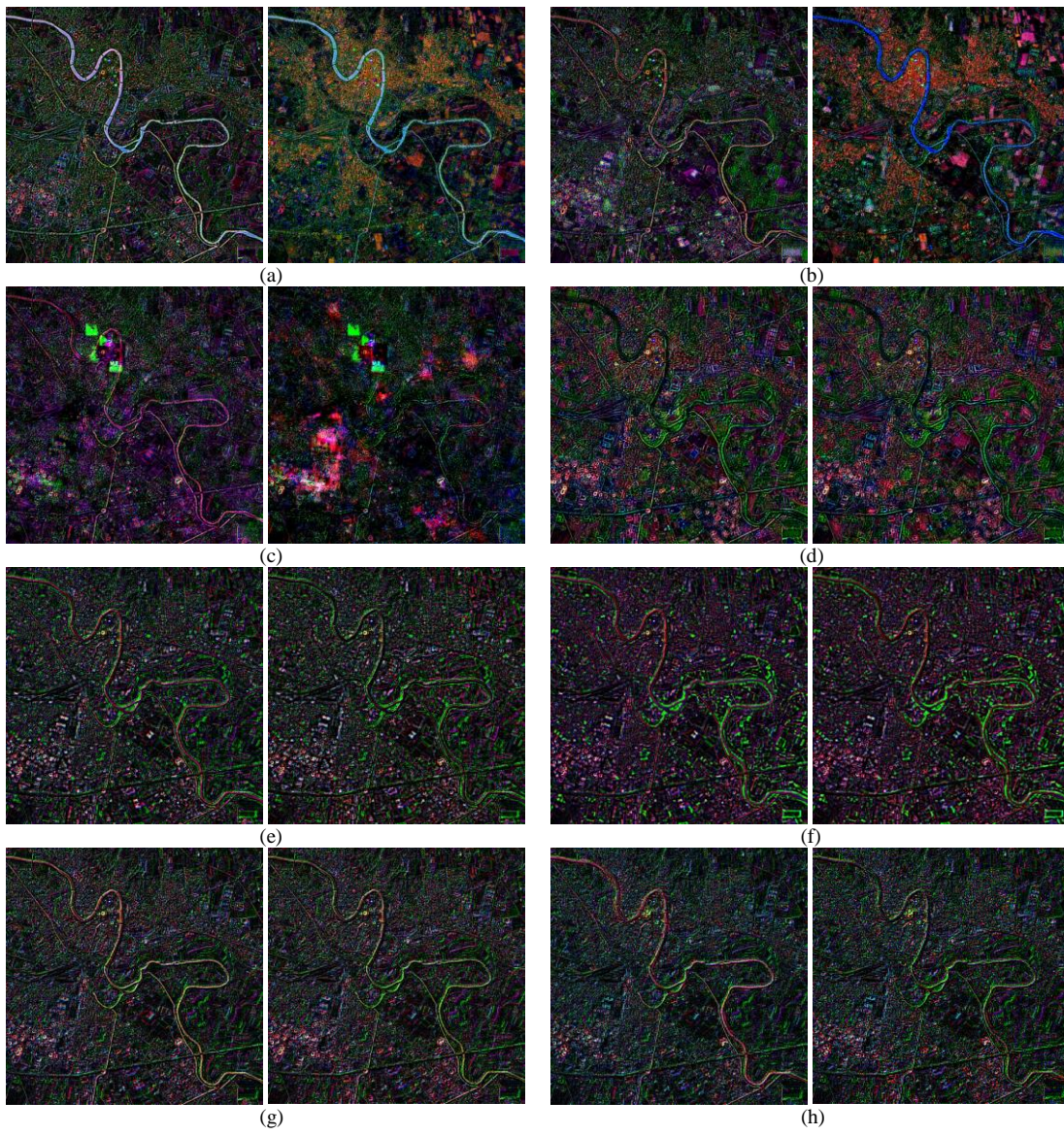
(a) (b) (c)
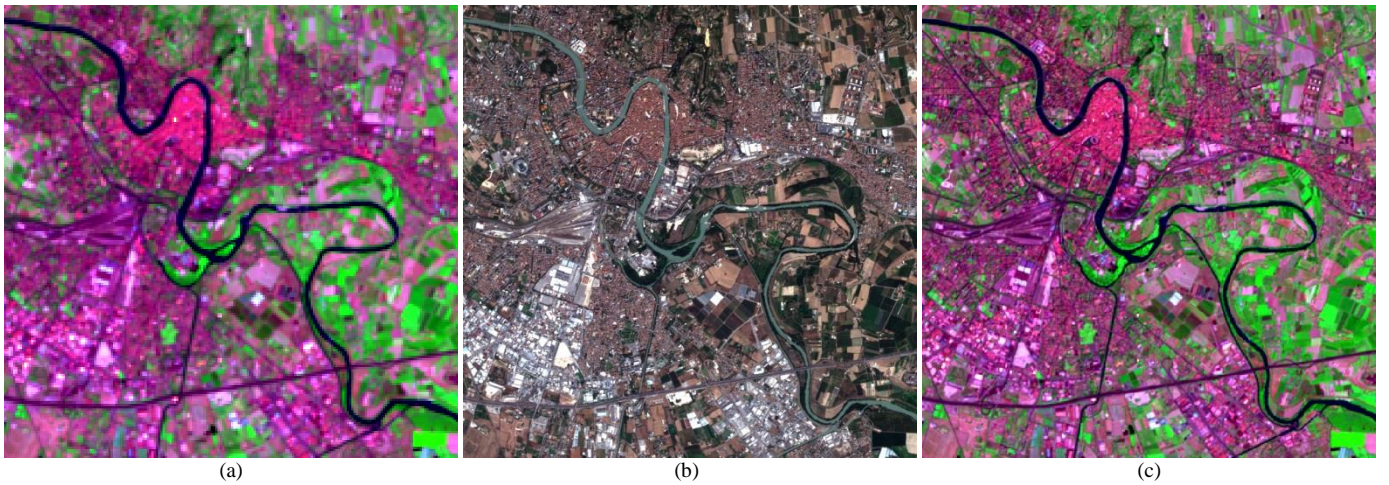
Fig. 5 Results for the Sentinel-2 dataset (bands 4, 3 and 2 as RGB for (b) and bands 12, 8a and 5 as RGB for (a) and (c)-(f)). (a) 40 m coarse image. (b) 20 m fine image. (c) 20 m reference. (d) 20 m polynomial interpolation result. (e) 20 m ATPK result. (f) 20 m ILGIF result.
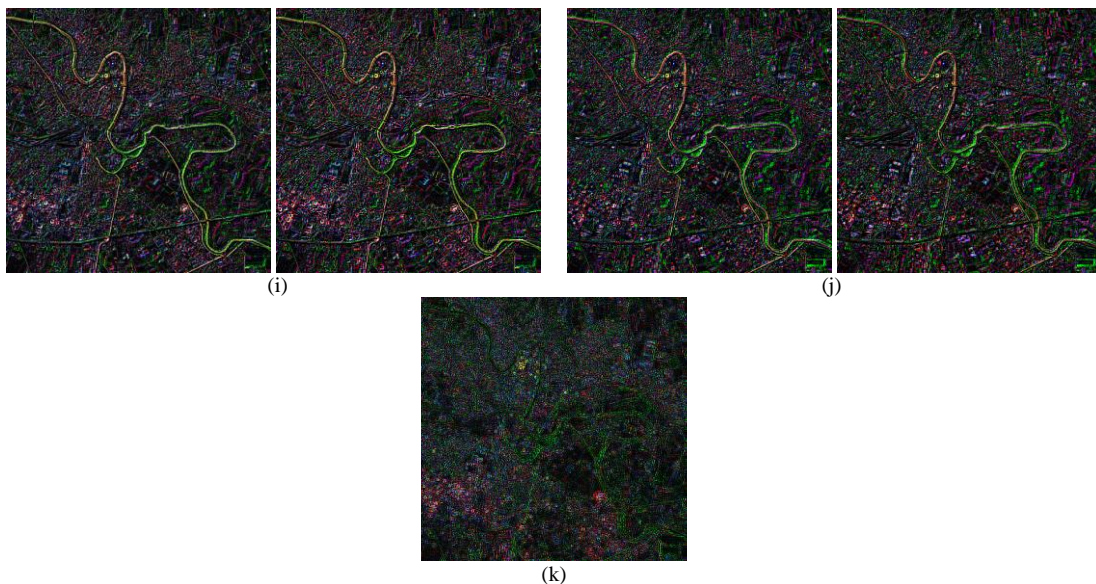
Fig. 6 Error maps for the different methods for the Sentinel-2 dataset (bands 12, 8a and 5 as RGB). For (a)-(j), the left and right are results for the selected and synthesized bands, respectively. (a) PRACS. (b) GSA. (c) GSA-CA. (d) BDSD. (e) ATWT. (f) AWLP. (g) MTF-GLP. (h) MTF-GLP-CBD. (i) MTF-GLP-HPM. (j) MF-HG. (k) ILGIF.

IV. DISCUSSION

Both ATPK and GWR are popular methods in spatial statistics. The proposed ILGIF method integrates them into a single framework for multi-resolution image fusion. ATPK is employed for initial downscaling, while GWR transforms the ILs from the fine bands covering the same area, but in other wavelengths, to that for the coarse band. Based on the important property of the IL (i.e., once upscaled to the coarse spatial resolution, it is exactly zero), it is concluded from (17) that the perfect coherence property of ILGIF is not influenced by the specific value of the weights in (13). This means that any weight can lead to a prediction with perfect coherence with the original coarse data. Such a property opens doors to more powerful alternatives to GWR for weight estimation.

In the experiments, when comparing the ILGIF predictions to the reference (ideal downscaling solution), there still exists gaps, which means IL still remains. The uncertainty in IL estimation in the ILGIF method may be ascribed to the inconsistency in terms of wavelength between the coarse band fine bands, as ILGIF treats the fine bands as training data and makes use of the ILs extracted from the fine bands. It would be worth developing more powerful models to relate the ILs from the fine bands to the coarse bands. Another possible choice for enhancement is to seek training data that fall in the same wavelength with the coarse band. As mentioned in the Introduction, such types of data may be challenging to provide as they need to be at the target fine spatial resolution and have a similar spatial pattern with the study area [6]. On the other hand, a large volume of such training data may be required to achieve as accurate a prediction as possible. This also motivates the development of more intelligent training schemes, such as that based on deep learning [40].

To reduce the smoothing effect in ATPK prediction and reproduce the variation at target fine spatial resolution, conditional simulation was developed in some literature [2], [41]. The idea of compensating ILs for the ATPK prediction in ILGIF is analogous to conditional simulation. However, they are substantially different. Specifically, for conditional simulation, an unconditional simulation at fine spatial resolution is produced first and then upscaled to match the spatial resolution of the input coarse data. The ATPK prediction for the simulated coarse data is compared to the available unconditional simulation and the difference (analogous to the IL defined in this paper) is finally added back to the ATPK prediction of the input coarse data [41]. Different unconditional simulation will lead to different prediction. Any prediction of conditional simulation has perfect coherence with the original coarse data. Admittedly, the conditional simulation scheme can increase the spatial variation of downscaling predictions, but this scheme is highly conditioned by the target spatial variation and the prediction always contains unstructured features, presenting as noise. This is because the unconditional simulation is derived from a random realization of white noise (zero-mean) without any spatial continuity. In this paper, however, IL is a fixed realization derived from fine bands which contain spatial continuity information highly related to the coarse bands (these bands were acquired over the same scene). ILGIF can, therefore, be viewed as a special case of conditional simulation, where the 'unconditional simulation' is actually a set of the available fine bands in different wavelengths.

Inheriting the advantages of ATPK, ILGIF accounts for the PSF and is suitable for any PSF. In the two experiments, we simulated coarse data based on the assumption of a Gaussian PSF, a filter widely used in remote sensing [42]-[44]. On the one hand, it should be noted that the sensor PSF in reality may be different from the Gaussian filter. For example, Tan *et al*. [45] claimed that the MODIS sensor has a scanning mirror which ensures that the shape has a directional component, and the sensor PSF was assumed to be triangular in the along-scan direction, but rectangular in the along-track direction. The characterization of the real PSF remains an open problem, and the most appropriate PSF model varies for different sensors. Specifically, the PSF depends on the used optics, the detector and the exploited scanning system and the electronics. Moreover,

it can vary over time due to the aging process [46]. As mentioned earlier, however, the implementation of ILGIF is not affected by the specific form of PSF, and any PSF can be readily used in ILGIF once it is known or estimated in advance.

It is necessary to use an accurate PSF in the ILGIF method. For example, for the Sentinel-2 dataset where the PSF was simulated with a Gaussian filter, when ILGIF was performed using a different square wave filter (i.e., the ideal PSF filter), the CC and UIQI of the prediction were 0.9361 and 0.9248, which are 0.054 and 0.065 smaller than those produced by the correct PSF. On the other hand, it should be stressed that when fusing images with different spatial resolutions, we are more interested in the PSF of the scale transformation than the PSF of the sensor (i.e., original measurement). It would be interesting to develop new methods to predict the mathematical formulation and corresponding parameters for the PSF in scale transformation. This is part of our ongoing research.

V. CONCLUSION

In this paper, based on the concept of IL, a new method called ILGIF is proposed for image fusion. ILGIF compensates the IL to the initial APTK prediction of the observed coarse image, where the IL is predicted using the ILs for fine spatial resolution bands acquired in other wavelengths. GWR is proposed to relate the two types of ILs and transform the ILs for the fine bands to the observed coarse band. ILGIF has the perfect coherence property and is suitable for pan-sharpening and fusion of multispectral and multi/hyperspectral images. Experiments on two datasets showed that ILGIF can produce more accurate results than six benchmark methods.

ACKNOWLEDGMENT

The authors would like to thank Dr. Gemine Vivone of the University of Salerno, Italy, for sharing the toolbox of pan-sharpening and for his kind discussion on their implementation considerations. The authors would also like to thank the Associate Editor and anonymous reviewers for their valuable and constructive comments which greatly improved the manuscript.

REFERENCES

[1] P. M. Atkinson, "Downscaling in remote sensing," *International Journal of Applied Earth Observation and Geoinformation*, vol. 22, pp. 106–114, 2013.

[2] P. Kyriakidis and E.-H. Yoo, "Geostatistical prediction and simulation of point values from areal data," *Geographical Analysis*, vol. 37, no. 2, pp. 124–151, 2005.

[3] P. C. Kyriakidis, "A geostatistical framework for area-to-point spatial interpolation," *Geographical Analysis*, vol. 36, no. 3, pp. 259–289, 2004.

[4] P. Goovaerts, "Kriging and semivariogram deconvolution in presence of irregular geographical units," *Mathematical Geosciences*, vol. 40, no. 1, pp. 101–128, 2008.

[5] A. Boucher and P. C. Kyriakidis, "Super-resolution land cover mapping with indicator geostatistics," *Remote Sensing of Environment*, vol. 104, no. 3, pp. 264–282, 2006.

[6] Q. Wang, P. M. Atkinson, W. Shi, "Indicator cokriging-based subpixel mapping without prior spatial structure information," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 1, pp. 309–323, 2015.

[7] M. Bevilacqua, A. Roumy, C. Guillemot, M.-L. Alberi Morel, "Single-image super-resolution via linear mapping of interpolated self-examples," *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 5334–5347, 2014.

[8] M. Drusch, et al., "Sentinel-2: ESA's optical high-resolution mission for GMES operational services," *Remote Sensing of Environment*, vol. 120, pp. 25–36, 2012.

[9] G. Vivone, L. Alparone, J. Chanussot, M. Dalla Mura, A. Garzelli, G. A. Licciardi, R. Restaino, and L. Wald, "A critical comparison among pansharpening algorithms," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 5, pp. 2565–2586, 2015.

[10] V. K. Shettigara, "A generalized component substitution technique for spatial enhancement of multispectral images using a higher resolution data set," *Photogrammetric Engineering and Remote Sensing*, vol. 58, no. 5, pp. 561–567, 1992.

[11] A. Garzelli, F. Nencini, and L. Capobianco, "Optimal MMSE pan sharpening of very high resolution multispectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 1, pp. 228–236, 2008.

[12] C. A. Laben, B. V. Brower, "Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening," U.S. Patent 6011875, 2000.

[13] B. Aiazzi, S. Baronti, and M. Selva, "Improving component substitution pansharpening through multivariate regression of MS+Pan data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 10, pp. 3230–3239, 2007.

[14] B. Aiazzi, S. Baronti, F. Lotti, M. Selva, "A comparison between global and context-adaptive pansharpening of multispectral images," *IEEE Geoscience and Remote Sensing Letters*, vol. 6, pp. 302–306, 2009.

[15] J. Choi, K. Yu, Y. Kim, "A new adaptive component-substitution based satellite image fusion by using partial replacement," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, pp. 295–309, 2011.

[16] P. S. Chavez Jr., S. C. Sides, and J. A. Anderson, "Comparison of three different methods to merge multiresolution and multispectral data: Landsat TM and SPOT panchromatic," *Photogrammetric Engineering and Remote Sensing*, vol. 57, no. 3, pp. 295–303, 1991.

[17] J. G. Liu, "Smoothing filter based intensity modulation: A spectral preserve image fusion technique for improving spatial details," *International Journal of Remote Sensing*, vol. 21, no. 18, pp. 3461–3472, 2000.

[18] G. Vivone, R. Restaino, M. Dalla Mura, G. Licciardi, J. Chanussot, "Contrast and error-based fusion schemes for multispectral image pan-sharpening," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, pp. 930–934, 2014.

[19] X. Otazu, M. González-Audícana, O. Fors, and J. Núñez, "Introduction of sensor spectral response into image fusion methods. Application to wavelet-based methods," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 10, pp. 2376–2385, 2005.

[20] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, M. Selva, "MTF-tailored multiscale fusion of high-resolution MS and Pan imagery," *Photogrammetric Engineering and Remote Sensing*, vol. 72, pp. 591–596, 2006.

[21] L. Alparone, L. Wald, J. Chanussot, C. Thomas, P. Gamba, and L. M. Bruce, "Comparison of Pansharpening Algorithms: Outcome of the 2006 GRS-S Data-Fusion Contest," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 10, pp. 3012–3021, 2007.

[22] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, "An MTF-based spectral distortion minimizing model for pan-sharpening of very high resolution multispectral images of urban areas," in *Proceedings of 2nd GRSS/ISPRS Joint Workshop Remote Sensing Data Fusion URBAN Areas*, 2003, pp. 90–94.

[23] Q. Wei, J. Bioucas-Dias, N. Dobigeon, and J. Tourneret, "Hyperspectral and multispectral image fusion based on a sparse representation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 7, pp. 3658–3668, 2015.

[24] Y. Wei, Q. Yuan, H. Shen, L. Zhang, "Boosting the accuracy of multispectral image pansharpening by learning a deep residual network," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 10, pp. 1795–1799, 2017.

[25] P. Ghamisi, B. Rasti, N. Yokoya, Q. Wang, B. Hofle, L. Bruzzone, F. Bovolo, M. Chi, K. Anders, R. Gloaguen, P. M. Atkinson, J. A. Benediktsson, "Multisource and multitemporal data fusion in remote sensing," *IEEE Geoscience and Remote Sensing Magazine*, vol. 7, no. 1, pp. 6–39, 2019.

[26] M. Selva, B. Aiazzi, F. Butera, L. Chiarantini, and S. Baronti, "Hyper-sharpening: A first approach on SIMGA data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 6, pp. 3008–3024, 2015.

[27] L. Wald, T. Ranchin, M. Mangolini, "Fusion of satellite images of different spatial resolutions: assessing the quality of resulting images,"

Photogrammetric Engineering and Remote Sensing, vol. 63, no. 6, pp. 691–699, 1997.

[28] G. M. Foody, D. S. Boyd, M. E. J. Cutler, "Predictive relations of tropical forest biomass from Landsat TM data and their transferability between regions," *Remote Sensing of Environment*, vol. 85, pp. 463– 474, 2003.

[29] Q. Wang, W. Shi, H. Zhang, "Class allocation for soft-then-hard subpixel mapping algorithms with adaptive visiting order of classes," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 9, pp. 1494–1498, 2014.

[30] C. Brunsdon, A. S. Fotheringham, M. E. Charlton, "Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity," *Geographical Analysis*, vol. 28, no. 4, pp. 281–298, 1996.

[31] A. S. Fotheringham, M. E. Charlton, C. Brunsdon, "Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis," *Environmental and Planning A*, vol. 30, pp. 1905–1927, 1998.

[32] W. Song, H. Jia, J. Huang, Y. Zhang, "A satellite-based geographically weighted regression model for regional PM2.5 estimation over the Pearl River Delta region in China," *Remote Sensing of Environment*, vol. 154, pp. 1–7, 2014.

[33] G.M. Foody, "Geographical weighting as a further refinement to regression modelling: An example focused on the NDVI–rainfall relationship," *Remote Sensing of Environment*, vol. 88, pp. 283–293, 2003.

[34] C. Zhang, W. Li, D. Civco, "Application of geographically weighted regression to fill gaps in SLC-off Landsat ETM+ satellite imagery," *International Journal of Remote Sensing*, vol. 35, no. 22, pp. 7650–7672, 2014.

[35] P. M. Atkinson, "Issues of uncertainty in super-resolution mapping and their implications for the design of an inter-comparison study," *International Journal of Remote Sensing*, vol. 30, no. 20, pp. 5293–5308, 2009.

[36] R. Restaino, G. Vivone, M. D. Mura, and J. Chanussot, "Fusion of multispectral and panchromatic images based on morphological operators," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 25, no. 6, pp. 2882–2895, 2016.

[37] A. Garzelli and F. Nencini, "Hypercomplex quality assessment of multi-/hyper-spectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 6, no. 4, pp. 662–665, 2009.

[38] L. Alparone, B. Aiazzi, S. Baronti, A. Garzelli, F. Nencini, and M. Selva, "Multispectral and panchromatic data fusion assessment without reference," *Photogrammetric Engineering and Remote Sensing*, vol. 74, no. 2, pp. 193–200, Feb. 2008.

[39] F. Palsson, J. R. Sveinsson, M. O. Ulfarsson, J. A. Benediktsson, "Quantitative quality evaluation of pansharpened imagery: Consistency versus synthesis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, pp. 1247–1259, 2016.

[40] L. Zhang, L. Zhang, B. Du, "Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art," *IEEE Geoscience and Remote Sensing Magazine*, vol. 4, no. 2, pp. 22–40, 2016.

[41] J. Wang, D. G. Brown, D. Hammerling, "Geostatistical inverse modeling for super-resolution mapping of continuous spatial processes," *Remote Sensing of Environment*, vol. 139, pp. 205–215, 2013.

[42] F. D. van der Meer, "Remote-sensing image analysis and geostatistics," *International Journal of Remote Sensing*, vol. 33, no. 18, pp. 5644–5676, 2012.

[43] B. N. Wenny, D. Helder, J. Hong, L. Leigh, K. J. Thome, D. Reuter, "Pre- and post-launch spatial quality of the Landsat 8 Thermal Infrared Sensor," *Remote Sensing*, vol. 7, pp. 1962–1980, 2015.

[44] Q. Wang, P. M. Atkinson, "The effect of the point spread function on sub-pixel mapping," *Remote Sensing of Environment*, vol. 193, pp. 127–137, 2017.

[45] B. Tan, et al., "The impact of gridding artifacts on the local spatial properties of MODIS data: Implications for validation, compositing, and band-to-band registration across resolutions," *Remote Sensing of Environment*, vol. 105, pp. 98–114, 2006.

[46] G. Vivone, M. Simoes, M. Dalla Mura, R. Restaino, J. M. Bioucas-Dias, G. A. Licciardi, J. Chanussot, "Pansharpening based on semiblind deconvolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, pp. 1997–2010, 2014.

Qunming Wang (M'15) received the Ph.D. degree from The Hong Kong Polytechnic University, Hong Kong, in 2015.

He is now a Professor in College of Surveying and Geo-Informatics, Tongji University, Shanghai, China. He was a Lecturer (Assistant Professor) in Lancaster Environment Centre, Lancaster University, UK from 2017 to 2018. His three-year PhD study was supported by the hyper-competitive Hong Kong PhD Fellowship and his PhD thesis was awarded as the outstanding thesis in the Faculty. His current research interests focus on remote sensing, image processing and geostatistics. He has published over 40 peer-reviewed articles in international journals such as *Remote Sensing of Environment*, *IEEE Transactions on Geoscience and Remote Sensing* and *ISPRS Journal of Photogrammetry and Remote Sensing*. He serves as a reviewer for 30 international journals, including most of the international journals in remote sensing. He is an Associate Editor for *Photogrammetric Engineering & Remote Sensing* and *Computers and Geosciences*.

Wenzhong Shi obtained the PhD degree from University of Osnabrück in Vechta, Germany, in 1994.

He is a Chair Professor in GIS and remote sensing, and the Head of Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University. His current research interests include GIS and remote sensing, uncertainty and spatial data quality control, image processing for high resolution satellite images. He has published over 160 SCI papers and 10 books. Prof. Shi received the State Natural Science Award from the State Council of China in 2007 and The Wang Zhizhuo Award from International Society for Photogrammetry and Remote Sensing in 2012.

Peter M. Atkinson received 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.

Peter is Dean of the Faculty of Science and Technology at Lancaster University. He was previously Professor of Geography at the University Southampton, where he is currently Visiting Professor. He is also Visiting Professor at Queen's University Belfast, UK and at the Chinese Academy of Sciences, Beijing, China. Peter previously held the Belle van Zuylen Chair at Utrecht University, the Netherlands and is recipient of the Peter Burrough Award of the International Spatial Accuracy Research Association. The main focus of Peter's 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 over 270 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.