A Comparative Analysis of Image Fusion Methods

Zhijun Wang, Djemel Ziou, Costas Armenakis, Deren Li, and Qingquan Li

Abstract—There are many image fusion methods that can be used to produce high-resolution mutlispectral images from a high-resolution panchromatic image and low-resolution mutlispectral images. Starting from the physical principle of image formation, this paper presents a comprehensive framework, the general image fusion (GIF) method, which makes it possible to categorize, compare, and evaluate the existing image fusion methods. Using the GIF method, it is shown that the pixel values of the high-resolution mutlispectral images are determined by the corresponding pixel values of the low-resolution panchromatic image, the approximation of the high-resolution panchromatic image at the low-resolution level. Many of the existing image fusion methods, including, but not limited to, intensity–hue–saturation, Brovey transform, principal component analysis, high-pass filtering, high-pass modulation, the à trous algorithm-based wavelet transform, and multiresolution analysis-based intensity modulation (MRAIM), are evaluated and found to be particular cases of the GIF method. The performance of each image fusion method is theoretically analyzed based on how the corresponding low-resolution panchromatic image is computed and how the modulation coefficients are set. An experiment based on IKONOS images shows that there is consistency between the theoretical analysis and the experimental results and that the MRAIM method synthesizes the images closest to those the corresponding multisensors would observe at the high-resolution level.

Index Terms—General image fusion (GIF) method, image formation principle, image fusion.

I. INTRODUCTION

THE CONCEPT of data fusion goes back to the 1950's and 1960's, with the search for practical methods of merging images from various sensors to provide a composite image which could be used to better identify natural and manmade objects. Terms such as merging, combination, synergy, integration, and several others that express more or less the same concept have since appeared in the literature. In the remote sensing community, the following definition has been adopted: "Data fusion is a formal framework in which are expressed means and tools for the alliance of data originating from different sources. It aims at obtaining information of greater quality; the exact definition of 'greater quality' will depend upon the application" [1].

Z. Wang and D. Ziou are with the Department of Computer Science, Université de Sherbrooke, Sherbrooke, QC J1K 2R1, Canada (e-mail: zhi.jun.wang@usherbrooke.ca; djemel.ziou@usherbrooke.ca).

C. Armenakis is with Natural Resources Canada, Centre for Topographic Information, Geomatics Canada, Ottawa, ON K1A 0E9, Canada (e-mail: armenaki@nrcan.gc.ca).

D. Li and Q. Li are with the National Laboratory for Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan, 430079, China (e-mail: drli@whu.edu.cn; qqli@whu.edu.cn).

In optical remote sensing, with physical and technological constraints, some satellite sensors supply the spectral bands needed to distinguish features spectrally but not spatially, while other satellite sensors supply the spatial resolution for distinguishing features spatially but not spectrally. For many applications, the combination of data from multiple sensors provides more comprehensive information. Several commercial earth observation satellites carry dual-resolution sensors of this kind, which provide high-resolution panchromatic images (HRPIs) and low-resolution multispectral images (LRMIs). For example, the first commercial high-resolution satellite, IKONOS, launched on September 24, 1999, produces 1-m HRPIs and 4-m LRMIs. This permits identification of objects approximately one meter in length (even less in some cases) on the earth's surface, using a satellite in outer space. It is particularly useful in urban areas because the characteristics of urban objects are determined not only by their spectra but also by their structure. It is therefore necessary and very useful to fuse HRPIs and LRMIs.

Many image fusion methods have been proposed for combining an HRPI with LRMIs. A detailed review on this issue was given by Pohl and Van Genderen [2]. Some methods, such as intensity–hue–saturation (IHS) [3]–[8], Brovey transform (BT) [9], [10], and principal component analysis (PCA) [10], [11], provide superior visual high-resolution multispectral images (HRMIs) but ignore the requirement of high-quality synthesis of spectral information. While these methods are useful for visual interpretation, high-quality synthesis of spectral information is very important for most remote sensing applications based on spectral signatures, such as lithology and soil and vegetation analysis [7]. Garguet-Duport *et al.* [12] has shown that the high-quality synthesis of spectral information is particularly well suited in the case of vegetation analysis. Wald *et al.* [13] suggests that the fused HRMIs should be as identical as possible to the real HRMIs the corresponding multispectral sensors would observe at the high-resolution level. A large amount of research has been done in terms of this constraint. The high-pass filtering (HPF) [14]–[18] and high-pass modulation (HPM) [16] methods and those of [19] and [20] have shown better performance in terms of the high-quality synthesis of spectral information. More recently, an underlying multiresolution analysis employing the discrete wavelet transform has been used in image fusion. It was found that multisensor image fusion is a tradeoff between the spectral information from an LRMI sensor and the spatial information from an HRPI sensor. With the wavelet transform fusion method, it is easy to control this tradeoff [10].

Currently used wavelet-based image fusion methods are mostly based on two computation algorithms: the Mallat algorithm [10], [18], [21]–[24] and the à trous algorithm [18], [25]–[28]. The Mallat algorithm-based dyadic wavelet transform (WT), which uses decimation, is not shift-invariant and

exhibits artifacts due to aliasing in the fused image [18], [26]. In contrast, the à trous algorithm-based dyadic wavelet transform (ATW) method, which does not use decimation, is shift-invariant, a characteristic that makes it particularly suitable for image fusion. Note that these methods are well suited to cases of image fusion where the resolution ratio between the LRMI and the HRPI is a power of two, although they are also used for ratios that are not powers of two, such as the fusion between the Systeme Pour l' Observation de la Terre (SPOT) panchromatic and Landsat Thematic Mapper multispectral bands [23]. However, some new methods have been proposed to deal better with cases whose ratios are not powers of two. Blanc *et al.* [29] used the ARSIS concept with iterated rational filter banks, and Shi *et al.* [30] used the Mallat algorithm and the M-band WT for the fusion of SPOT panchromatic and Landsat Thematic Mapper multispectral bands. Moreover, Wang [31] used the multiresolution analysis-based intensity modulation (MRAIM) method based on the à trous algorithm and the M-band WT (the main strategy behind this approach will be introduced in a later section), and Aiazzi *et al.* [18] extended the Laplacian pyramid (LP) to the generalized LP (GLP) method to deal with cases whose ratios are arbitrary integers.

The existing image fusion methods can be classified into several groups. Schowengerdt [16] classified them into spectral domain techniques, spatial domain techniques, and scale space techniques. Ranchin and Wald [24] classified them into three groups: projection and substitution methods, relative spectral contribution methods, and those relevant to the ARSIS concept. It was found that many of the existing image fusion methods, such as the HPF, WT, and ATW methods, can be accommodated within the ARSIS concept. Tu *et al.* [8] also performed a mathematical evaluation and found that the PCA, BT, and ATW methods can be accommodated as IHS-like image fusion methods. In this vein, the main objective of this paper is to propose a comprehensive framework, the general image fusion (GIF) method, which makes it possible to categorize, compare, and evaluate the existing image fusion methods.

The paper is structured in six sections. The following section explains the GIF method. Section III introduces the mathematical models of several existing image fusion methods such as IHS, BT, HPF, HPM, PCA, ATW, and MRAIM. In Section IV, the relations between the GIF method and the existing methods are examined and the performance of the existing methods is theoretically analyzed. In Section V, experiments conducted based on IKONOS images are presented with their results. Finally, our conclusions are given in Section VI.

II. PHYSICAL PRINCIPLE OF IMAGE FORMATION AND THE GIF METHOD

Optical remote sensing sensors all operate on the same principle of optical radiation transfer, image formation, and photon detection. The visible (0.4–0.7 μm), near infrared (0.7–1.1 μm), and shortwave infrared (1.1–1.35 μm, 1.4–1.8 μm, and 2–2.5 μm) regions are the primarily solar-reflective spectral range, because within this range the energy reflected by the earth's surface exceeds that emitted by the earth itself [16]. Reflective panchromatic and multispectral band sensors see the reflected solar energy and produce panchromatic and multispectral images. Different materials have different
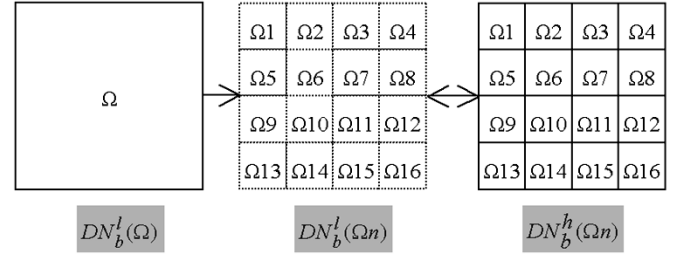

Fig. 1. Relationship between the low-resolution pixel and the corresponding high-resolution pixels.

spectral reflection and absorption features and can therefore be mapped in terms of different pixel digital numbers (DNs) in these images [7].

There are generally three significant components in the upwelling at-sensor radiation, which are: 1) the unscattered, surface-reflected radiation; 2) the downscattered, surface-reflected skylight; and 3) the upscattered path radiance. Let us assume that atmospheric correction has been done so that the atmospheric effects of skylight and path radiance have been removed from the images [16], [32]–[34]. With the exception of cast shadows, the spectral irradiance at the top of the earth's atmosphere E^0 is assumed to be a constant across the scene and the atmospheric transmittance (both the solar path τ_s and the view path τ_v) is assumed to be the same in each band. These constants are combined with the gain factor of the sensor calibration coefficients of each band into a single constant k_b, which is band-dependent and resolution-independent. Under the assumption that the terrain is a Lambertian surface, the DN in band b is approximately a linear function of at-sensor radiance and also a linear function of earth surface reflectance ρ_b. Let γ denote the solar incidence angle to the terrain surface at pixel (x, y) and O_b denote the offset factor of the sensor calibration coefficients. We can then write [16]

$$\mathrm{DN}_b(x,y) = k_b \rho_b(x,y) \cos[\gamma(x,y)] + O_b. \qquad (1)$$

The DN value results from the product of two spatially varying terms, the reflectance and the cosine factor. Either may be considered to spatially modulate the other [16]. However, the cosine factor is strictly terrain-dependent and causes the DN values of the pixels of different bands to vary in the same way.

The GIF method is derived from the simplified image formation model. Without loss of generality, let us take IKONOS sensors as an example [now the O_b is zero in (1)]. Suppose that an IKONOS sensor sees a terrain space $\Omega(x, y)$ at band b (panchromatic, blue, green, red, or NIR) with 1- and 4-m resolutions, respectively. The terrain space Ω (called the pixel below) is 4 m $\times$ 4 m, corresponding to one IFOV of a 4-m low-resolution pixel. We get one low-resolution pixel value at 4-m resolution and 16 high-resolution pixel values at 1-m resolution (see Fig. 1). The low-resolution pixel value is denoted as $\mathrm{DN}_b^l(\Omega(x,y))$, and the 16 high-resolution pixel values are denoted as $\{\mathrm{DN}_b^h(\Omega_n(x,y)), n = 1, \ldots, 16\}$, where the subscript b denotes the band of the image and the superscripts h and l denote high and low resolution. $\Omega_n(x,y)$ denotes the low-resolution pixel $\Omega(x,y)$ partitioned into 16 high-resolution pixels, where $n = 1, \ldots, 16$. Let us consider a space corresponding to

a high-resolution pixel as a pure unit surface on which the reflectance ρ and the solar incidence angle γ are uniform. From (1), we have

$$\mathrm{DN}_b^h(\Omega_n) = k_b \rho_b(\Omega_n) \cos[\gamma(\Omega_n)], \qquad n = 1, \ldots, 16. \quad (2)$$

Since each low-resolution pixel value (or radiance) can be treated as a weighted average of high-resolution pixel values (or radiances) over the corresponding space, provided that only the uniformly weighted case is considered for the sake of simplicity, we have

$$
\begin{aligned}
\mathrm{DN}_b^l(\Omega) &= \frac{1}{16} \sum_1^{16} \mathrm{DN}_b^h(\Omega_n) \\
&= \frac{k_b}{16} \sum_1^{16} \rho_b(\Omega_n) \cos[\gamma(\Omega_n)]. \quad (3)
\end{aligned}
$$

However, for pixel-level image fusion, it is necessary to sample the low-resolution image and the high-resolution image at the same pixel size. The low-resolution mother pixel $\mathrm{DN}_b^l(\Omega)$ is thus replaced with 16 child pixels at 1-m pixel size, whose pixel values may be considered to be the same as the $\mathrm{DN}_b^l(\Omega)$, for the sake of simplification, and denoted as $\{\mathrm{DN}_b^l(\Omega_n), n = 1, \ldots, 16\}$. Then the one-to-one relationship between the pixels of the high-resolution image and the pixels of the low-resolution image is established at 1-m pixel size (see Fig. 1). If we divide the corresponding pixel values of the same band high-resolution and low-resolution images, then k_b is cancelled. We obtain

$$
\begin{aligned}
\frac{\mathrm{DN}_b^h(\Omega_n)}{\mathrm{DN}_b^l(\Omega_n)} &= \frac{\mathrm{DN}_b^h(\Omega_n)}{\mathrm{DN}_b^l(\Omega)} \\
&= \frac{16 \rho_b(\Omega_n) \cos[\gamma(\Omega_n)]}{\rho_b(\Omega_1) \cos[\gamma(\Omega_1)] + \cdots + \rho_b(\Omega_{16}) \cos[\gamma(\Omega_{16})]}. \quad (4)
\end{aligned}
$$

Within a low-resolution image, pixels can be classified as pure pixels and mixed pixels. Pure pixels correspond to the interior pixels (homogeneous regions), which originate from the same type of high-resolution pixels. The spatial difference between the high-resolution pixels and the corresponding pure low-resolution pixels is caused by the cosine factor. Mixed pixels correspond to the boundary pixels (heterogeneous regions), which originate from different types of high-resolution pixels. The spatial difference between the high-resolution pixels and the corresponding mixed low-resolution pixels is caused by both reflectance and cosine variations.

For pure low-resolution pixels, we have

$$\rho_b(\Omega_1) = \rho_b(\Omega_2) = \cdots = \rho_b(\Omega_{16}) = \rho_b(\Omega). \quad (5)$$

Equation (4) thus becomes

$$\frac{\mathrm{DN}_b^h(\Omega_n)}{\mathrm{DN}_b^l(\Omega_n)} = \frac{16 \cos[\gamma(\Omega_n)]}{\cos[\gamma(\Omega_1)] + \cdots + \cos[\gamma(\Omega_{16})]}. \quad (6)$$

Equation (6) shows that the ratio between $\mathrm{DN}_b^h(\Omega_n)$ and $\mathrm{DN}_b^l(\Omega_n)$ is controlled by the cosine factor γ and is independent of the sensor band. It is worth mentioning that independence of the sensor band is only approximately true. Perfect independence of the sensor band is never achieved because image

formation in reality is far more complex than the simplified image formation model.

If PAN represents a panchromatic band and MS a multispectral band, then from (6), for $b = \mathrm{PAN}$ and $b = \mathrm{MS}$, we have

$$\frac{\mathrm{DN}_{\mathrm{PAN}}^h(\Omega_n)}{\mathrm{DN}_{\mathrm{PAN}}^l(\Omega_n)} = \frac{\mathrm{DN}_{\mathrm{MS}}^h(\Omega_n)}{\mathrm{DN}_{\mathrm{MS}}^l(\Omega_n)}. \quad (7)$$

Manipulating (7) yields the equivalent expression

$$\mathrm{DN}_{\mathrm{MS}}^h(\Omega_n) = \frac{\mathrm{DN}_{\mathrm{PAN}}^h(\Omega_n)}{\mathrm{DN}_{\mathrm{PAN}}^l(\Omega_n)} \mathrm{DN}_{\mathrm{MS}}^l(\Omega_n). \quad (8)$$

In practice, $\mathrm{DN}_{\mathrm{PAN}}^h(\Omega_n)$, the pixel values of the HRPI, and $\mathrm{DN}_{\mathrm{MS}}^l(\Omega_n)$, the corresponding pixel values of the upsampled LRMI, are known. $\mathrm{DN}_{\mathrm{PAN}}^l(\Omega_n)$, the pixel values of the low-resolution panchromatic image (LRPI), and $\mathrm{DN}_{\mathrm{MS}}^h(\Omega_n)$, the pixel values of the HRMI, are unknown. However, as long as $\mathrm{DN}_{\mathrm{PAN}}^l(\Omega_n)$ is approximated, $\mathrm{DN}_{\mathrm{MS}}^h(\Omega_n)$ can be solved for. The mathematical model itself is the same as the HPM method and has also appeared in many other papers, such as those of [19], [20], andd [33]. However, different methods are used to compute $\mathrm{DN}_{\mathrm{PAN}}^l(\Omega_n)$ in those papers.

Since (8) can lead to an ill-posed situation if $\mathrm{DN}_{\mathrm{PAN}}^l(\Omega_n) = 0$, introducing two intermediate parameters w and α in (8) gives

$$\mathrm{DN}_{\mathrm{MS}}^h(\Omega_n) = \mathrm{DN}_{\mathrm{MS}}^l(\Omega_n) + \alpha w \quad (9)$$

where $w = \mathrm{DN}_{\mathrm{PAN}}^h(\Omega_n) - \mathrm{DN}_{\mathrm{PAN}}^l(\Omega_n), \alpha = (\mathrm{DN}_{\mathrm{MS}}^l(\Omega_n))/(\mathrm{DN}_{\mathrm{PAN}}^l(\Omega_n))$ when $\mathrm{DN}_{\mathrm{PAN}}^l(\Omega_n) \neq 0$ and α is set to 1 when $\mathrm{DN}_{\mathrm{PAN}}^l(\Omega_n) = 0$. Parameter w is the signal difference between the HRPI and the LRPI. It represents the detail information between the high- and low-resolution levels. Parameter α is the modulation coefficient for the detail information. It determines how the detail information of the HRPI is injected into the LRMIs.

For mixed low-resolution pixels, (7) does not always hold. One solution in this case is to subdivide these pixels into several homogeneous sections according to the information from high-resolution pixels; (7) will then hold true within each homogeneous section. However, it is hard to un-mix the spectral properties ρ_b of these pixels when the HRPI has only one spectral band. For the sake of simplicity, we assume that (7) holds true for these pixels as well, so that spatial information caused by variations in reflectance is restored in the same way as that resulting from cosine variation.

Equation (9) is considered as the mathematical model of the GIF method. It shows that the pixel values of the HRMIs are determined by the corresponding pixel values of the LRPI, or, say, by the corresponding detail information of the panchromatic band and the modulation coefficients. The performance of the fusion process is therefore determined by how the LRPI is computed and how the modulation coefficients are set.

It is also worth mentioning that, while the GIF method is derived based on the IKONOS sensor, it is also applicable to other sensors where the calibration coefficient offset in (1) is zero. For those sensors whose offset is not equal to zero, the DN must be calibrated to the radiance or, at least, the offset must be calibrated from the DN in order to make the GIF method hold true.

III. PRINCIPLES OF SEVERAL EXISTING IMAGE FUSION METHODS

In order to better demonstrate that existing image fusion methods can be accommodated by the framework of the GIF method, the mathematical models of a number of methods are sorted into one group of typical exemplars rather than investigating all of the existing methods exhaustively. The representative methods dealt with here are IHS, BT, HPF, HPM, PCA, ATW, and MRAIM. Although some of these methods have different applications or extended models (for instance, the IHS method has both cylindrical and spherical coordinates), only the standard model of each is investigated. In the literature, different notations are used for the various methods; for the sake of convenience, the same notation is used for all methods in this paper. For purposes of theoretical comparison, the original mathematical models of the methods are adjusted to fit the form of (9). For all methods, it is assumed that the images are geometrically registered, and the LRMIs are upsampled to the pixel size of the HRPI.

A. IHS Transform

The IHS technique is a standard procedure in image fusion, with the major limitation that only three bands are involved [3]–[6]. Originally, it was based on the RGB true color space. It offers the advantage that the separate channels outline certain color properties, namely intensity (I), hue (H), and saturation (S). This specific color space is often chosen because the visual cognitive system of human beings tends to treat these three components as roughly orthogonal perceptual axes. However, in remote sensing, arbitrary bands are usually assigned to the RGB channels to produce false color composites for display purposes only. The IHS technique usually comprises four steps: 1) transform the red, green, and blue (RGB) channels (corresponding to three multispectral bands) to IHS components; 2) match the histogram of the panchromatic image with the intensity component; 3) replace the intensity component with the stretched panchromatic image; and 4) inverse-transform IHS channels to RGB channels. The resultant color composite will then have a higher spatial resolution in terms of topographic texture information. The forward transform of this method is described by (10), while (11) describes the backward transformation. The original transformation matrix of the IHS transformation is orthogonal

$$
\begin{bmatrix} DN_{PAN}^l \\ V1 \\ V2 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{-1}{\sqrt{6}} & \frac{-1}{\sqrt{6}} & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} & \frac{-1}{\sqrt{6}} & 0 \end{bmatrix} \begin{bmatrix} DN_{MS1}^l \\ DN_{MS2}^l \\ DN_{MS3}^l \end{bmatrix} \quad (10)
$$

and $H = \tan^{-1}[V2/V1], S = \sqrt{V1^2 + V2^2}$

$$
\begin{bmatrix} DN_{MS1}^h \\ DN_{MS2}^h \\ DN_{MS3}^h \end{bmatrix} = \begin{bmatrix} 1 & \frac{-1}{\sqrt{6}} & \frac{3}{\sqrt{6}} \\ 1 & \frac{-1}{\sqrt{6}} & \frac{-3}{\sqrt{6}} \\ 1 & \frac{2}{\sqrt{6}} & 0 \end{bmatrix} \begin{bmatrix} DN_{PAN}^{h'} \\ V1 \\ V2 \end{bmatrix}. \quad (11)
$$

Since the forward and backward transformations are both linear, replacing V1 and V2 in (11) by V1 and V2 from (12) yields the mathematical model of the generalized IHS method

$$
\begin{bmatrix} DN_{MS1}^h \\ DN_{MS2}^h \\ DN_{MS3}^h \end{bmatrix} = \begin{bmatrix} DN_{MS1}^l \\ DN_{MS2}^l \\ DN_{MS3}^l \end{bmatrix} + \left(DN_{PAN}^{h'} - DN_{PAN}^l \right) \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \quad (12)
$$

where $DN_{PAN}^l = (1/3)(DN_{MS1}^l + DN_{MS2}^l + DN_{MS3}^l)$ and $DN_{PAN}^{h'}$ is DN_{PAN}^h, stretched to have same mean and variance as DN_{PAN}^l.

B. Brovey Transform

The BT is based on the chromaticity transform [7], [9], [10]. It is a simple method for combining data from different sensors, with the limitation that only three bands are involved. Its purpose is to normalize the three multispectral bands used for RGB display and to multiply the result by any other desired data to add the intensity or brightness component to the image. The original mathematical model is the same as that of (8). For the sake of comparison, it is given as

$$
\begin{bmatrix} DN_{MS1}^h \\ DN_{MS2}^h \\ DN_{MS3}^h \end{bmatrix} = \begin{bmatrix} DN_{MS1}^l \\ DN_{MS2}^l \\ DN_{MS3}^l \end{bmatrix} + \left(DN_{PAN}^h - DN_{PAN}^l \right) \begin{bmatrix} \frac{DN_{MS1}^l}{DN_{PAN}^l} \\ \frac{DN_{MS2}^l}{DN_{PAN}^l} \\ \frac{DN_{MS3}^l}{DN_{PAN}^l} \end{bmatrix} \quad (13)
$$

where $DN_{PAN}^l = (1/3)(DN_{MS1}^l + DN_{MS2}^l + DN_{MS3}^l)$.

C. High-Pass Filtering

The principle of HPF is to add the high-frequency information from the HRPI to the LRMIs to get the HRMIs [16]–[18]. The high-frequency information is computed by filtering the HRPI with a high-pass filter or taking the original HRPI and subtracting the LRPI, which is the low-pass filtered HRPI. This method preserves a high percentage of the spectral characteristics, since the spatial information is associated with the high-frequency information of the HRMIs, which is from the HRPI, and the spectral information is associated with the low-frequency information of the HRMIs, which is from the LRMIs. The mathematical model is

$$
DN_{MS}^h = DN_{MS}^l + \left(DN_{PAN}^h - DN_{PAN}^l \right) \quad (14)
$$

where $DN_{PAN}^l = DN_{PAN}^h * h_0$ and h_0 is a low-pass filter such as a boxcar filter. When boxcar filters are used, the filter length is crucial and must match the resolution ratio of the HRPI and LRMIs. A 3×3 boxcar filter is suitable for $1:2$ fusion only, since the frequency response should have -6-dB cutoff (halved amplitude) at $f_n = 0.25$, where f_n is the spatial frequency normalized to the sampling frequency f_s. For $1:4$ fusion, a 5×5 boxcar filter with cutoff frequency at roughly 0.125 must be used. Fig. 2 illustrates the frequency responses of 3×3, 5×5, and 7×7 boxcar filters. It shows that a smooth transition band is accompanied by a large ripple outside the pass-band [18].

D. High-Pass Modulation

The principle of HPM is to transfer the high-frequency information of the HRMI to the LRMIs, with modulation coefficients, which equal the ratio between the LRMIs and the LRPI. The LRPI is obtained by low-pass filtering the HRPI. The equivalent mathematical model is [16]

$$
DN_{MS}^h = DN_{MS}^l + \left(DN_{PAN}^h - DN_{PAN}^l \right) \frac{DN_{MS}^l}{DN_{PAN}^l} \quad (15)
$$

where $DN_{PAN}^l = DN_{PAN}^h * h_0$ and h_0 is the same low-pass filter as used in the HPF method.
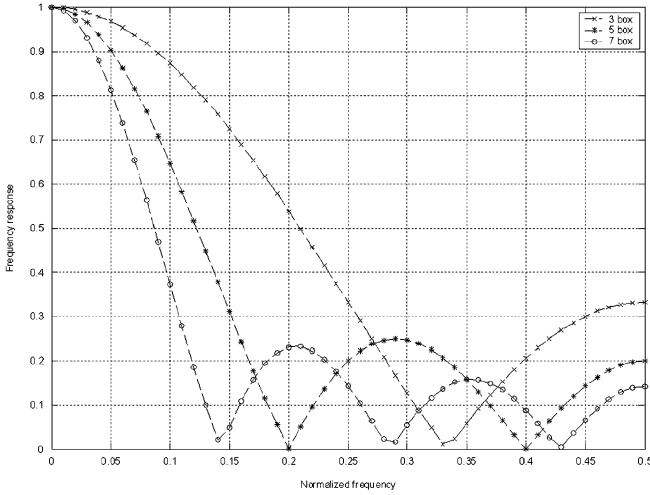
Fig. 2. Frequency responses of boxcar filters for the HPF method.

E. Principal Component Analysis

The PCA method is similar to the IHS method, with the main advantage that an arbitrary number of bands can be used [10], [11]. The input LRMIs are first transformed into the same number of uncorrelated principal components. The first principal component image contains the information that is common to all the bands used as input to PCA, while the spectral information that is unique to any of the bands is mapped to the other components. Then, similar to the IHS method, the first principal component (PC1) is replaced by the HRPI, which is first stretched to have the same mean and variance as PC1. As a last step, the HRMIs are determined by performing the inverse PCA transform.

The mathematical models of the forward and backward processes are represented by (16) and (17). The transformation matrix v contains the eigenvectors, ordered with respect to their eigenvalues. It is orthogonal and determined either from the covariance matrix or the correlation matrix of the input LRMIs. PCA performed using the covariance matrix is referred to as unstandardized PCA, while PCA performed using the correlation matrix is referred to as standardized PCA.

$$\begin{bmatrix} PC1 \\ PC2 \\ \cdots \\ PCn \end{bmatrix} = \begin{bmatrix} v11 & v21 & \cdots & vn1 \\ v12 & v22 & \cdots & vn2 \\ \cdots & \cdots & \cdots & \cdots \\ v1n & v2n & \cdots & vnn \end{bmatrix} \begin{bmatrix} DN_{MS1}^l \\ DN_{MS2}^l \\ \cdots \\ DN_{MSn}^l \end{bmatrix} \quad (16)$$

where the transformation matrix

$$v = \begin{bmatrix} v11 & v12 & \cdots & v1n \\ v21 & v22 & \cdots & v2n \\ \cdots & \cdots & \cdots & \cdots \\ vn1 & vn2 & \cdots & vnn \end{bmatrix}.$$

$$\begin{bmatrix} DN_{MS1}^h \\ DN_{MS2}^h \\ \cdots \\ DN_{MSn}^h \end{bmatrix} = \begin{bmatrix} v11 & v12 & \cdots & v1n \\ v21 & v22 & \cdots & v2n \\ \cdots & \cdots & \cdots & \cdots \\ vn1 & vn2 & \cdots & vnn \end{bmatrix} \begin{bmatrix} DN_{PAN}^{h'} \\ PC2 \\ \cdots \\ PCn \end{bmatrix}. \quad (17)$$

Similar to the IHS method, (16) and (17) can be merged as follows:

$$\begin{bmatrix} DN_{MS1}^h \\ DN_{MS2}^h \\ \cdots \\ DN_{MSn}^h \end{bmatrix} = \begin{bmatrix} DN_{MS1}^l \\ DN_{MS2}^l \\ \cdots \\ DN_{MSn}^l \end{bmatrix} + (DN_{PAN}^{h'} - DN_{PAN}^l) \begin{bmatrix} v11 \\ v21 \\ \cdots \\ vn1 \end{bmatrix} \quad (18)$$

where $DN_{PAN}^l = PC1$ and $DN_{PAN}^{h'}$ is DN_{PAN}^h, stretched to have same mean and variance as PC1.

F. À Trous Algorithm-Based Wavelet Transform

The term "à trous" was originally introduced in [35]. It is based on the undecimated dyadic wavelet transform and is particularly suitable for signal processing since it is isotropic and shift-invariant and does not create artifacts when used in image processing [25], [27]. Its application to image fusion is reported in [18] and [26]. Given an image p, the sequence of its approximations by multiresolution decomposition is $F_1(p) = p_1, F_2(p) = p_2, \ldots$. The wavelet planes w_j are computed as the differences between two consecutive approximations [27]

$$w_j = p_{j-1} - p_j, \qquad j = 1, \ldots, r, \ p_0 = p. \quad (19)$$

The reconstruction formula can be written as

$$p = p_r + \sum_{j=1}^{r} w_j \quad (20)$$

where r refers to the decomposition level. To construct the sequence, the à trous algorithm performs successive convolutions with a filter h_0 obtained from the scaling function. The use of a B3 cubic spline yields a dyadic low-pass scaling function such as $h_0 = (1/16)[1, 4, 6, 4, 1]$ in one dimension. There are several possible ways of using the à trous algorithm in image fusion [18], [26], [28]. Here, for the sake of clarity, we follow the additive method (AWRGB) described in [26]. The ATW method is given by

$$DN_{MS}^h = DN_{MS}^l + (DN_{PAN}^h - DN_{PAN}^l) \quad (21)$$

where $DN_{PAN}^l = p_r$, p_r is the approximation of the panchromatic image at decomposition level r, defined according to the resolution ratio. For $1:2$ fusion, r is set to 1; for $1:4$ fusion, r is set to 2; and so on.

G. Multiresolution Analysis-Based Intensity Modulation

MRAIM was proposed by Wang [31]. It follows the GIF method, with the major advantage that it can be used for the fusion case in which the ratio is an arbitrary integer M, with a very simple scheme. The mathematical model is

$$DN_{MS}^h = DN_{MS}^l + (DN_{PAN}^h - DN_{PAN}^l) \frac{DN_{MS}^l}{DN_{PAN}^l} \quad (22)$$

where $DN_{PAN}^l = DN_{PAN}^h * h_0$ and, for $1:M$ fusion (where M is an arbitrary integer), h_0 is an M-band low-pass filter that satisfies the à trous condition: $h_{Mk} = \delta(k)/\sqrt{M}$. It is driven by
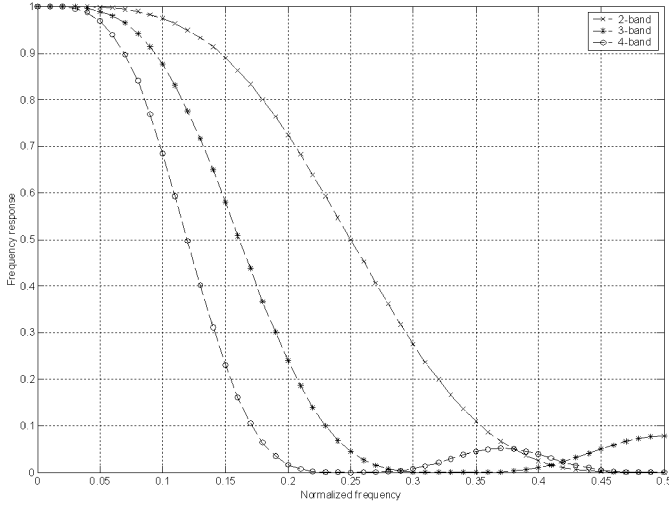
Fig. 3.   Frequency responses of two-band, three-band, and four-band à trous low-pass filters ($K = 2$).

TABLE I
RELATIONS BETWEEN THE GIF METHOD AND SEVERAL
EXISTING IMAGE FUSION METHODS

	$a = c$	$\alpha = \dfrac{DN_{low}(\lambda_{MS})}{DN_{low}(\lambda_{pan})}$
$DN_{low}(\lambda_{pan}) = \frac{1}{3}\sum_{i=1}^{3} DN_{low}(\lambda_{MSi})$	IHS	BT
$DN_{low}(\lambda_{pan}) = \sum_{i=1}^{3} a_i DN_{low}(\lambda_{MSi})$	PCA	*
$DN_{low}(\lambda_{pan}) = DN_{high}(\lambda_{pan}) * h_0$ h_0 is a boxcar filter	HPF	HPM
$DN_{low}(\lambda_{pan}) = DN_{high}(\lambda_{pan}) * h_0$ h_0 is a Gaussian-like low-pass filter	ATW	*
$DN_{low}(\lambda_{pan}) = DN_{high}(\lambda_{pan}) * h_0$ h_0 is an M-band à trous low-pass filter	*	MRAIM

* No corresponding method has yet been reported in the literature
c: constant value

the squares of K-regularity M-band scale filters (in the convolution sense).[1] K-regularity is equivalent to saying that all polynomials of degree ($K - 1$) are contained in the approximation and the polynomials of higher degree (equal to and larger than K) are contained in the detail. This, coupled with the compact support characteristic of the scaling filters, implies that K-regularity scaling filters can be used to capture local polynomial behaviors. For $K = 2$, Fig. 3 gives the frequency responses of two-band, three-band, and four-band low-pass filters as examples. It shows that the frequency responses have cutoff frequency at $f_n = 0.5/2$ for a two-band low-pass filter, $0.5/3$ for a three-band low-pass filter, and $0.5/4$ for a four-band low-pass filter, where f_n is the spatial frequency f normalized by the sampling frequency f_s. These low-pass filters can be used for $1:2$, $1:3$, and $1:4$ fusion, respectively. A detailed presentation of the MRAIM method has been submitted to the *ISPRS Journal of Photogrammetry and Remote Sensing* and is now under review.

IV. RELATIONS BETWEEN THE GIF METHOD AND EXISTING IMAGE FUSION METHODS

In the preceding section, the mathematical models of the seven methods were expressed similarly. After comparing with the mathematical model of the GIF method, we found that these methods can be regarded as particular cases of the GIF method.

In the GIF method, a pixel value of the HRMI is determined as a function of its corresponding LRPI and the modulation coefficient. In practical implementation, the LRPI is treated as a weighted average over the corresponding HRPI, or simulated by a weighted average over the LRMIs. The modulation coefficient is simplified as a constant value or kept as originally defined. Thus, following the GIF method, image fusion methods can be classified into four classes: 1) $DN_{PAN}^l = f(MS_1, \ldots, MS_n)$, and $\alpha = c$; 2) $DN_{PAN}^l = f(MS_1, \ldots, MS_n), \alpha_i = (DN_{MSi}^l)/(DN_{PAN}^l)$; 3) $DN_{PAN}^l = f(DN_{PAN}^h)$, and $\alpha = c$; and 4) $DN_{PAN}^l = f(DN_{PAN}^h)$ and $\alpha_i = (DN_{MSi}^l)/(DN_{PAN}^l)$. Classes 1 and 2

correspond to the spectral domain technique, classes 3 and 4 to the spatial domain technique, and when the function f in classes 3 and 4 is based on the multiresolution analysis technique, then classes 3 and 4 correspond to the scale domain technique in [16] and [36].

The IHS, BT, HPF, HPM, PCA, ATW, and MRAIM methods follow the GIF method. The IHS, BT, and PCA methods use a linear combination of the LRMIs to compute the LRPIs, with different coefficients. The HPF, HPM, ATW, and MRAIM methods compute the LRPIs by low-pass filtering the original HRPI with different filters. The BT, HPM, and MRAIM methods use the modulation coefficients as the ratios between the LRMIs and the LRPI, whereas the IHS, HPF, ATW, and PCA methods simplify the modulation coefficients to constant values for all pixels of each band. It is obvious that the IHS and PCA methods belong to class 1, the BT method belongs to class 2, the HPF and ATW methods belong to class 3, and the HPM and MRAIM methods belong to class 4. These relations are summarized in Table I. It is obvious that many other methods, such as that of [20], also follow the GIF method.

The performance of each image fusion method is determined by two factors: how the LRPI is computed and how the modulation coefficients are defined. For the methods in classes 1 and 2, since DN_{PAN}^l is approximated from the LRMIs, it usually has a weak correlation with DN_{PAN}^h. Note that the correlation between images of different resolutions has no intrinsic significance, but it can be used to compare the behavior of the different methods [26]. The process of histogram matching, which stretches the HRPI to have the same mean and variation as DN_{PAN}^l, may be applied to narrow the difference and ensure that the energy of the detail information w is small. However, histogram matching is a global process: the distortion of spectral characteristics is globally reduced but not removed, generating color distortion in the fused image. For the methods in classes 3 and 4, the LRPI is from a low-pass filtering process; no histogram matching is needed in this case and usually the results have less spectral distortion. For the methods in classes 1 and 3 where $\alpha = c$, the reflectance differences between the panchromatic and multispectral bands have not been taken into consideration, and each band of the HRMIs involves the addition of the same component w to the original LRMIs; the fused images bias the color of the pixel toward the gray [26]. The methods in classes 2 and 4, where $\alpha_i = (DN_{MSi}^l)/(DN_{PAN}^l)$,

---

[1]The construction of K-regularity M-band scale filters is explained in more detail at http://www.ee.umanitoba.ca/~ferens/wavelets.html.

are all ratio methods. These methods can preserve the ratios between the respective bands and give more emphasis to slight signature variations. Their aim is to maintain the radiometric integrity of the data while increasing the spatial resolution. It is worth mentioning that several other models, such as Model 2 and the RWM model of the ARSIS concept [28], are also used to set the modulation coefficients. However, we only investigate the above two models from the perspective of the GIF method.

The performance of the existing methods is therefore analyzed based on the above theoretical evaluation. The performance of the IHS, BT, and PCA methods relies on the correlation between the panchromatic image and the multispectral images. Somewhat better performance can be expected when the available datasets have strong correlation. Otherwise, it is difficult or impossible to achieve optimal spectral and spatial results simultaneously, even if the process of histogram matching is applied during the fusion process. The BT method will get even worse results, since no histogram matching process is applied to narrow the difference between the LRPI and the HRPI. The HPF, HPM, ATW, and MRAIM methods follow a different approach: the LRPI is obtained by low-pass filtering the HRPI. Therefore less spectral distortion will be introduced in the results. In the case of the HPF method, a boxcar filter is usually used as the low-pass filter. The filter length is crucial and must match the ratio of the HRPI and LRMIs. However, the ripple in the frequency response will have some negative impact. The similar HPM method will give slightly better performance because the color of the pixels is not biased toward the gray, provided that the low-pass filter is chosen as in the HPF method. The ATW method represents a step forward from HPF: the low-pass filter, a cubic spline scaling function introduced in [27], is roughly a Gaussian filter without the ripple. Finally, the MRAIM method exactly follows the GIF method and differs from the HPM method only in the low-pass filter. As discussed earlier, the frequency responses of M-band low-pass filters have halved amplitude at $f_n = 0.5/M$, enabling computation of the LRPI with arbitrary scale ratio equal to $1 : M$. It is thus easy to control how much high-frequency information is transferred to the LRMIs. In particular, M-band low-pass filters are more flexible than other low-pass filters, especially when the scale ratio is not a power of two.

V. EXPERIMENTS AND RESULTS

In order to validate the theoretical analysis, the performance of the methods discussed above was further evaluated by experimentation. The IKONOS-2 panchromatic band (525.8–928.5 μm) of the 1-m resolution HRPI and the red (631.9–697.7 μm), green (506.4–595.0 μm), blue (444.7–516.0 μm) and NIR (757.3–852.7 μm) bands of the 4-m resolution LRMIs were used in this experiment. The images, covering an area of the city of Sherbrooke, QC, Canada, were captured on May 20, 2001. The size of the HRPI is $688 * 1100$ and the size of the original LRMIs is $172 * 275$, but this is upsampled to $688 * 1100$ by bicubic interpolation. The pairs of images were geometrically registered to each other. IKONOS data was collected at 11 bits per pixel (2048 gray tones). This means that there is more definition in the gray scale values and the viewer can see more detail in the image. In order to benefit from this additional information, the processing and evaluation

Fig. 4.   Original HRPI (panchromatic band).

Fig. 5.   Original LRMIs (RGB) (resampled at 1-m pixel size).

TABLE II
CORRELATION COEFFICIENTS FOR THE IKONOS HRPI (RESAMPLED
AT 4-m PIXEL SIZE) AND THE ORIGINAL LRMIs

Red	Green	Blue	NIR
0.4432	0.4479	0.3452	0.6409

were entirely based on the original 11-bit data and the data was converted to eight-bit for display purposes only. Fig. 4 shows the HRPI. To save space, only the natural color image of the original LRMIs (red–green–blue combination) resampled at 1-m pixel size is shown in Fig. 5. The NIR band is not shown but was processed and numerically evaluated. The study area is composed of various features such as cars, buildings, trees, lawn, etc., ranging in size from less than 1 m up to 100 m. It is obvious that the HRPI has better spatial resolution than the LRMIs and more detail can be seen from the HRPI. Table II gives the correlation coefficients (CCs) between the HRPI (downsampled to 4-m pixel size) and the original LRMIs, which show that the CC of the NIR band is higher than the CCs of other bands, indicating that the IKONOS NIR band is very important to the IKONOS PAN band.

The IHS and BT methods can only handle three bands. In order to evaluate the NIR band as well, we selected the red–green–blue combination for true natural color and the NIR–red–green combination for false color. In comparing with other methods, however, only the NIR band of the second combination was used with the red, green, and blue bands
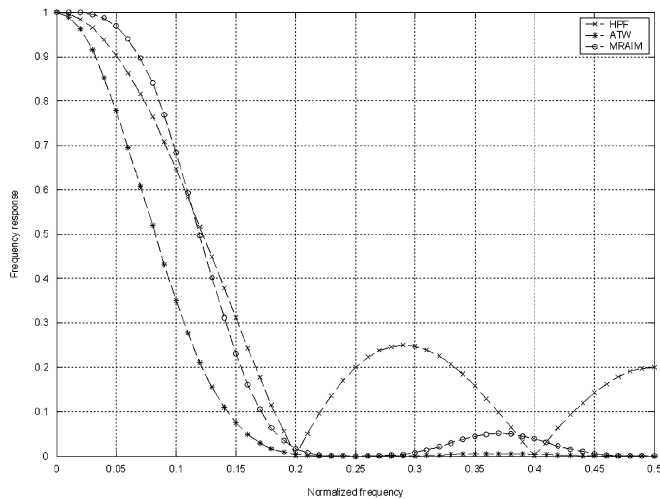
Fig. 6. Frequency responses of the low-pass filters used by the HPF, ATW, and MRAIM methods for 1 : 4 fusion.

Fig. 8. Result of the BT method.

Fig. 9. Result of the PCA method.

Fig. 7. Result of the IHS method.

Fig. 10. Result of the HPF method.

of the first combination. The resolution ratio between the IKONOS HRPI and the LRMIs is 1 : 4. Therefore, in the HPF and HPM methods, a 5×5 boxcar filter was used. In the ATW method, the Gaussian-like low-pass filter is suitable for 1 : 2 fusion only, so two-level filtering was necessary. In the MRAIM method, the four-band low-pass filter with regularity factor $K = 2$ was used. The frequency responses of these three filters are displayed in Fig. 6, which shows that the HPF/HPM and MRAIM methods have almost the same cutoff frequency at $f_n = 0.125$ but the HPF/HPM method has greater ripple outside the pass-band. The frequency response of the ATW method is the product of a two-level process and has its cutoff frequency roughly at 0.07, which is lower than the 0.125 of the HPF/HPM and MRAIM methods. Hence, more high-frequency information from the HRPI will unexpectedly be transferred to the LRMIs by the ATW method.

The fused results of the IHS, BT, PCA, HPF, HPM, ATW, and MRAIM methods are displayed as Figs. 7–13, respectively. Since the results are too large to be assessed together, for better evaluation, Fig. 14 shows subscenes from the original natural color composite and the corresponding results together.

Our purpose is to simulate the HRMIs as the real HRMIs (RHRMIs) the corresponding multispectral sensors would observe at the high-resolution level. So the performance of each

fusion method should be evaluated in terms of the quality of synthesis of both spatial and spectral information, which means that the HRMIs should be as identical as possible to the RHRMIs. Note that the techniques, which perform the most radiometrically accurate sharpening, are usually not those that produce the synthesized images with which give the best visual impression. Radiometric accuracy should therefore be used as the principal criterion. However, we do not actually have the RHRMIs to compare with, so evaluating the performance of the different methods is not easy.

Fig. 11. Result of the HPM method.

Fig. 12. Result of the ATW method.

Fig. 13. Result of the MRAIM method.

To solve the problem that no RHRMIs exist in actual practice, a protocol proposed in [13], and also used in [26], degrades the original HRPI and LRMIs to an inferior resolution level, and then treats the original LRMIs as the real high-resolution multispectral images to compare with the synthesized images. The protocol assesses quality by checking for three properties: 1) any synthesized image once degraded to its original resolution should be as identical as possible to the original image; 2) any synthesized image should be as identical as possible to the image that the corresponding sensor would observe with the

highest spatial resolution; and 3) the multispectral set of synthesized images should be as identical as possible to the multispectral set of images that the corresponding sensor would observe with the highest spatial resolution. This is based on the assumption that if properties 2) and 3) are satisfied at an inferior level, they are also satisfied at the original resolution level. Although studies have demonstrated that the quality of the synthesized images at the original resolution level cannot be predicted from the assessments at the inferior level, it nonetheless seems reasonable to assume that the quality of the synthesized images at the original resolution level is close to the quality of the synthesized images at the inferior level [37], so properties 2) and 3) can still be used to compare the behavior of the various methods. In our study, quality was evaluated both qualitatively and quantitatively. Visual inspection was used for the qualitative assessment since visual inspection is a subjective but effective tool for showing the major advantages and disadvantages of a method. Wald's protocol was used for the quantitative evaluation.

Visual inspection provided an overall impression of the detail information and the similarity of the original and resultant images. By visually comparing the spatial quality of all the resultant images (Figs. 7–13) with that of the original image (Fig. 5), it is obvious that the spatial resolutions of the resultant images are higher than that of the original image. Some small features such as cars, which were not interpretable in the original image, can be identified individually in each of the resultant images. Trees, buildings, and highways are much sharper in the resultant images than in the original image. It is easy to see this effect in Fig. 14. This means that all of the methods can improve spatial quality via the fusion process. In Figs. 7–9, the IHS, BT, and PCA methods produce significant color distortion with respect to the original image. In Figs. 10 and 12, the HPF and ATW methods produce color distortion in instances such as lawn and dark forested areas. In Fig. 11, the HPM method also exhibits slight color distortion, as in the bright builtup area, for instance, but is better than the HPF and ATW methods (see also Fig. 14). This may be because of the large ripple outside its pass-band in the frequency response of its low-pass filter. Nevertheless the MRAIM method looks better than all of the other methods in terms of the quality of spectral information. The additive HPF and ATW methods look sharper than the modulated HPM and MRAIM methods; however, this is probably due to over-enhancement along the edge area because these additive methods have not considered the differences in high-frequency information between the panchromatic band and the multispectral bands, so this should not be considered as a merit of the HPF and ATW methods. The quality of spectral information is the principal criterion. In Fig. 14, it can be seen that the MRAIM method also gives better spatial quality than the HPM method. Overall, it is obvious by visual inspection that the MRAIM method gives the synthesized result closest to what the corresponding multisensors would observe at the high-resolution level.

In addition to the visual analysis, the performance of each method was further quantitatively analyzed by checking the three properties. The correlation coefficient (CC) is the most popular similarity metric in image fusion. However, CC is insensitive to a constant gain and bias between two images and does not allow subtle discrimination of possible fusion artifacts
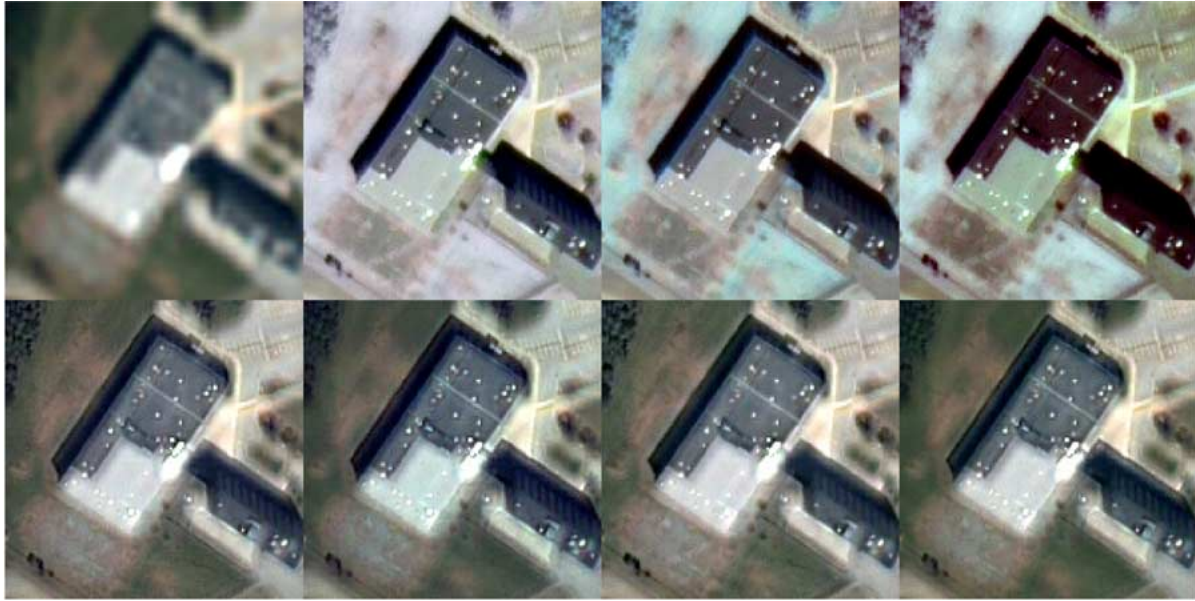
Fig. 14. Subscenes of the original LRMIs and the fused resulting HRMIs by different methods (double zoom). (Left to right sequence, row by row) Original LRMIs, IHS, BT, PCA, HPF, HPM, ATW, and MRAIM.

TABLE III
UIQI MEASUREMENT OF SIMILARITY BETWEEN THE DEGRADED FUSED
IMAGE AND THE ORIGINAL IMAGE AT 4-m RESOLUTION LEVEL

	Red	Green	Blue	NIR
IHS	0.4381	0.4404	0.1893	0.9283
BT	0.6081	0.3806	0.1610	0.9225
PCA	0.3830	0.3922	0.1424	0.9195
HPF	0.9931	0.9926	0.9886	0.9945
HPM	0.9952	0.9932	0.9905	0.9932
ATW	0.9806	0.9807	0.9685	0.9909
MRAIM	*0.9962*	*0.9949*	*0.9931*	*0.9947*

TABLE IV
CORRELATION BETWEEN THE IKONOS HRPI AND THE CORRESPONDING
LRPIs BY DIFFERENT METHODS (COMPUTED AT 1-m PIXEL SIZE)

IHS & BT	PCA	HPF & HPM	ATW	MRAIM
0.3822 (R, G, B)	0.2549	0.9262	0.8920	0.9337
0.7529 (NIR, R, G)				

Both the IHS and BT methods are limited to three bands. The 0.3822 is computed for the Red, Green, and Blue band combination. The 0.7529 is computed for the NIR, Red, and Green combination.

[18]. Recently, a universal image quality index (UIQI) [38], [39] has been used to measure the similarity between two images. In this experiment, we used the UIQI to measure similarity. The UIQI is designed by modeling any image distortion as a combination of three factors: loss of correlation, radiometric distortion, and contrast distortion. It is defined as follows:

$$Q = \frac{\sigma_{AB}}{\sigma_A \sigma_B} \cdot \frac{2\mu_A \mu_B}{\mu_A^2 + \mu_B^2} \cdot \frac{2\sigma_A \sigma_B}{\sigma_A^2 + \sigma_B^2}. \tag{23}$$

The first component is the correlation coefficient for A and B. The second component measures how close the mean gray levels of A and B is, while the third measures the similarity between the contrasts of A and B. The dynamic range is $[-1, 1]$. If two images are identical, the similarity is maximal and equals 1.

In order to test the first property, the synthesized HRMIs are spatially degraded to the resolution level of the original LRMIs (4 m) by bicubic interpolation. Then the UIQIs are computed between the degraded HRMIs (DHRMIs) and the original LRMIs at the 4–m resolution level. Table III shows that the UIQIs of the IHS, BT, and PCA methods are somewhat lower than the UIQIs of the other methods. The UIQI of the MRAIM method is higher than the UIQIs of the other low-pass filtering methods. This may be because the M-band low-pass filter has better low-pass characteristics than other low-pass filters, such as having its cutoff frequency at $f_n = 0.5/M$ and relatively small ripple outside the pass-band (see Fig. 6). This means that the IHS, BT,

TABLE V
UIQIs FOR THE RESULTANT IMAGES AND THE ORIGINAL
LRMIs AT 4 m. (FUSION AT THE INFERIOR LEVEL)

	Red	Green	Blue	NIR	Average
IHS	0.4654	0.4528	0.2272	0.8468	0.4981
BT	0.5659	0.3887	0.1941	0.8429	0.4979
PCA	0.4389	0.4329	0.1928	0.7457	0.4526
HPF	0.7930	0.7958	0.7280	0.8471	0.7910
HPM	0.8337	0.8043	0.7521	0.8064	0.7991
ATW	0.7905	0.7922	0.7189	*0.8503*	0.7880
MRAIM	*0.8779*	*0.8134*	*0.7644*	0.8165	*0.8181*

and PCA methods distort the spectral characteristics more than other methods and the MRAIM method synthesizes the spectral information better. Table IV shows the correlation between the IKONOS HRPI and the corresponding LRPIs by different methods. It can be seen from Tables III and IV that the degree of similarity between the HRPI and the LRPI corresponds to the degree of spectral distortion of each method. The lower the similarity between the HRPI and the LRPI, the higher the spectral distortion, and vice versa. In all bands, the MRAIM method gives the highest score. All of the methods give higher scores in the NIR band. This is because the PAN and NIR bands are more highly correlated. The PAN band includes the most important information from the NIR band.

In order to test the second and third properties, the HRPI is degraded to 4-m pixel size and the LRMIs are degraded to 16-m pixel size by bicubic interpolation. Then each image fusion method is applied to the degraded data to yield the resultant

image at the 4-m resolution level, after which the resultant images are evaluated against the original LRMIs at the 4-m resolution level. Table V shows that the MRAIM method gives the highest scores for the red, green, and blue bands. All of the methods yield high scores for the NIR band. However, at the NIR band, the MRAIM method does not give the best score, which is not consistent with the results from the test on the first property. This may be because all the methods provide good results in the NIR band, so the difference is very small, while the spatial degradation process will influence the final result differently for different fusion methods. Since in-depth evaluation of the degradation process is beyond the scope of this paper, we can only say that although we assume the influence due to interpolation is very small, it is not negligible in this case. This also confirms Wald's observation [37] that the quality of the synthesized images at the original resolution level cannot be predicted from the assessments at the inferior level.

The second property deals with the bands individually while the third property evaluates the quality of all the bands together. We use the average of the UIQIs of LRMIs as the score. The ranks of different methods in Table V correspond to the ranks of the respective methods in Table IV. It also shows that the MRAIM method gives the highest score.

By combining the visual inspection results and the quantitative results, it can be seen that the experimental results are in conformity with the theoretical analysis. The IHS, BT, and PCA methods produce considerable spectral distortion; the HPF, HPM, and ATW methods produce slight spectral distortion. The MRAIM method produces the images closest to those the corresponding multisensors would observe at the high-resolution level.

VI. CONCLUSION

This paper proposes a framework, the GIF method. Under different assumptions on how the LRPI is computed and how the modulation coefficients are set, many existing image fusion methods, including, but not limited to, IHS, BT, HPF, HPM, PCA, ATW, and MRAIM, are shown to be particular cases of the GIF method. The performance of each method is determined by two factors: how the LRPI is computed and how the modulation coefficients are defined. If the LRPI is approximated from the LRMIs, it usually has a weak correlation with the HRPI, leading to color distortion in the fused image. If the LRPI is a low-pass filtered HRPI, it usually shows less spectral distortion. If the modulation coefficient is set as a constant value, the reflectance differences between the panchromatic bands and the multispectral bands are not taken into consideration, and the fused images bias the color of the pixel toward the gray. Methods in which the modulation coefficients are set following the GIF method can preserve the ratios between the respective bands, give more emphasis to slight signature variations, and maintain the radiometric integrity of the data while increasing spatial resolution.

By combination of the visual inspection results and the quantitative results, it is possible to see that the experimental results are in conformity with the theoretical analysis and that the MRAIM method produces the synthesized images closest to those the corresponding multisensors would observe at the high-resolution level.

REFERENCES

[1] L. Wald, "Some terms of reference in data fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1190–1193, May 1999.

[2] C. Pohl and J. L. Van Genderen, "Multi-sensor image fusion in remote sensing: Concepts, methods, and applications," *Int. J. Remote Sens.*, vol. 19, no. 5, pp. 823–854, 1998.

[3] P. S. Chavez and J. A. Bowell, "Comparison of the spectral information content of Landsat thematic mapper and SPOT for three different sites in the Phoenix, Arizona region," *Photogramm. Eng. Remote Sens.*, vol. 54, no. 12, pp. 1699–1708, 1988.

[4] W. J. Carper, T. M. Lillesand, and R. W. Kiefer, "The use of Intensity-Hue-Saturation transform for merging SPOT panchromatic and multi-spectral image data," *Photogramm. Eng. Remote Sens.*, vol. 56, no. 4, pp. 459–467, 1990.

[5] K. Edwards and P. A. Davis, "The use of Intensity-Hue-Saturation transformation for producing color shaded-relief images," *Photogramm. Eng. Remote Sens.*, vol. 60, no. 11, pp. 1369–1374, 1994.

[6] E. M. Schetselaar, "Fusion by the IHS transform: Should we use cylindrical or spherical coordinates?," *Int. J. Remote Sens.*, vol. 19, no. 4, pp. 759–765, 1998.

[7] 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.

[8] T. M. Tu, S. C. Su, H. C. Shyu, and P. S. Huang, "A new look at IHS-like image fusion methods," *Inf. Fusion*, vol. 2, no. 3, pp. 177–186, 2001.

[9] A. R. Gillespie, A. B. Kahle, and R. E. Walker, "Color enhancement of highly correlated images—II. Channel ratio and 'chromaticity' transformation techniques," *Remote Sens. Environ.*, vol. 22, pp. 343–365, 1987.

[10] J. Zhou, D. L. Civco, and J. A. Silander, "A wavelet transform method to merge Landsat TM and SPOT panchromatic data," *Int. J. Remote Sens.*, vol. 19, no. 4, pp. 743–757, 1998.

[11] P. S. Chavez and A. Y. Kwarteng, "Extracting spectral contrast in Landsat Thematic Mapper image data using selective principle component analysis," *Photogramm. Eng. Remote Sens.*, vol. 55, no. 3, pp. 339–348, 1989.

[12] B. Garguet-Duport, J. Girel, J.-M. Chassery, and G. Pautou, "The use of multi-resolution analysis and wavelets transform for merging SPOT panchromatic and multi-spectral image data," *Photogramm. Remote Sens.*, vol. 62, no. 9, pp. 1057–1066, 1996.

[13] L. Wald, T. Ranchin, and M. Mangolini, "Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images," *Photogramm. Eng. Remote Sens.*, vol. 63, no. 6, pp. 691–699, 1997.

[14] R. A. Schowengerdt, "Reconstruction of multi-spatial, multi-spectral image data using spatial frequency content," *Photogramm. Eng. Remote Sens.*, vol. 46, no. 10, pp. 1325–1334, 1980.

[15] P. S. Chavez, S. C. Sides, and J. A. Anderson, "Comparison of three different methods to merge multi-resolution and multi-spectral data: Landsat TM and SPOT panchromatic," *Photogramm. Eng. Remote Sens.*, vol. 57, no. 3, pp. 295–303, 1991.

[16] R. A. Schowengerdt, *Remote Sensing: Models and Methods for Image Processing*, 2nd ed. Orlando, FL: Academic, 1997.

[17] S. de Béthune, F. Muller, and J. P. Donnay, "Fusion of multi-spectral and panchromatic images by local mean and variance matching filtering techniques," *Fusion of Earth Data*, Jan. 28–30, 1998.

[18] B. Aiazzi, L. Alparone, S. Baronti, and A. Garzelli, "Context-driven fusion of high spatial and spectral resolution images based on oversampled multi-resolution analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 10, pp. 2300–2312, Oct. 2002.

[19] J. G. Liu and J. M. Moore, "Pixel block intensity modulation: Adding spatial detail to TM band 6 thermal imagery," *Int. J. Remote Sens.*, vol. 19, no. 13, pp. 2477–2491, 1998.

[20] Y. Zhang, "A new merging method and its spectral and spatial effects," *Int. J. Remote Sens.*, vol. 20, no. 10, pp. 2003–2014, 1999.

[21] S. Mallat, "A theory for multi-resolution signal: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, Jul. 1989.

[22] D. A. Yocky, "Image merging and data fusion by means of the discrete two-dimensional wavelet transform," *J. Opt. Soc. Amer. A*, vol. 12, no. 9, pp. 1834–1841, 1995.

[23] ——, "Multiresolution wavelet decomposition image merger of Landsat Thematic Mapper and SPOT Panchromatic data," *Photogramm. Eng. Remote Sens.*, vol. 62, no. 9, pp. 1067–1074, 1996.

[24] T. Ranchin and L. Wald, "Fusion of high spatial and spectral resolution images: The ARSIS concept and its implementation," *Photogramm. Eng. Remote Sens.*, vol. 66, no. 1, pp. 49–61, 2000.

[25] M. J. Shensa, "The discrete wavelet transform: Wedding the *à Trous* and Mallat algorithms," *IEEE Trans. Signal Process.*, vol. 40, no. 10, pp. 2464–2482, Oct. 1992.

[26] J. Núñez, X. Otazu, O. Fors, A. Prades, V. Palà, and R. Arbiol, "Multiresolution-based image fusion with additive wavelet decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1204–1211, May 1999.

[27] F. Murtagh and J. L. Starck, "Image processing through multiscale analysis and measurement noise modeling," *Stat. Comput.*, vol. 10, no. 2, pp. 95–103, 2000.

[28] T. Ranchin, B. Aiazzi, L. Alparone, S. Baronti, and L. Wald, "Image fusion—The ARSIS concept and some successful implementation schemes," *ISPRS J. Photogramm. Remote Sens.*, vol. 58, pp. 4–18, 2003.

[29] P. Blanc, T. Blu, T. Ranchin, L. Wald, and R. Aloisi, "Using iterated rational filter banks within the ARSIS concept for producing 10 m Landsat multi-spectral images," *Int. J. Remote Sens.*, vol. 19, no. 12, pp. 2331–2343, 1998.

[30] W. Z. Shi, C. Q. Zhu, C. Y. Zhu, and X. M. Yang, "Multi-band wavelet for fusing SPOT panchromatic and multi-spectral images," *Photogramm. Eng. Remote Sens.*, vol. 69, no. 5, pp. 513–520, 2003.

[31] Z. J. Wang, "Wavelet transform-based multi-sensor remote sensing image fusion," Ph.D. thesis, Wuhan University, Wuhan, 2000. In Chinese.

[32] P. S. Chavez, "Radiometric calibration of Landsat thematic mapper multi-spectral images," *Photogramm. Eng. Remote Sens.*, vol. 55, pp. 1285–1294, 1989.

[33] D. Tanré *et al.*, "Description of a computer code to simulate the satellite signal in the solar spectrum: The 5S code," *Int. J. Remote Sens.*, vol. 11, pp. 659–668, 1990.

[34] D. Pradines, "Improving SPOT image size and multi-spectral resolution," in *Proc. SPIE Conf. Earth Remote Sensing Using the Landsat Thematic Mapper and SPOT Sensor Systems*, vol. 660, Tyrol, Austria, Apr. 15–17, 1986, pp. 98–102.

[35] P. Dutilleux, "An implementation of the "algorithme à trous" to compute the wavelet transform," in *Wavelets: Time-Frequency Methods and Phase Space*, J. M. Combes, A. Grossman, and Ph. Tchamitchian, Eds. Berlin, Germany: Springer-Verlag, 1989, pp. 298–304.

[36] B. Zhukov, D. Oertel, F. Lanzl, and G. Reinhäckel, "Unmixing-based multi-sensor multi-resolution image fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1212–1226, May 1999.

[37] L. Wald, *Data Fusion: Definitions and Architectures—Fusion of Images of Different Spatial Resolutions*. Paris, France: Les Presses de l'Ecole des Mines, 2002. .

[38] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002.

[39] L. Bentabet, S. Jodouin, D. Ziou, and J. Vaillancourt, "Road vectors update using SAR imagery: A snake-based method," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 8, pp. 1785–1803, Aug. 2003.

Djemel Ziou received the B.Eng. degree in computer sciences from Université d'Annaba, Annaba, Algeria, in 1984, and the Ph.D. degree in computer sciences from the Institut National Polytechnique de Lorraine, Lorraine, France, in 1991.

From 1987 to 1993, he served as Lecturer in several universities in France. During the same period, he was a Researcher in the Centre de Recherche en Informatique de Nancy (CRIN), Nancy, France, and the Institut National de Recherche en Informatique (INRIA), Paris, France. He is currently a Full Professor in the Department of Computer Science, University of Sherbrooke, Sherbrooke, QC, Canada. His research interests include image processing, computer vision, and pattern recognition. He heads the Centre MOIVRE and the consortium CoRIMedia for the automatic management of multimedia content, which he founded. He has published about 140 papers on feature extraction, statistical modeling, content-based image retrieval, image modeling, and sensor and data fusion.

Dr. Ziou has served on numerous conference committees as member or chair.

Costas Armenakis received the Dipl. Ing. in rural and surveying engineering from the National Technical University, Athens, Greece, and the M.Sc.E. and Ph.D. degrees in surveying engineering from the University of New Brunswick, Fredericton, NS, Canada, in 1979, 1983, and 1988, respectively.

Currently he is a Research Scientist with the Centre for Topographic Information, Natural Resources Canada, Ottawa, ON, and works in the fields of digital photogrammetry and remote sensing for topographic mapping applications. His research interests are in the areas of rapid revision of topographic data involving change detection, feature extraction, and classification from imagery. He is also involved with data integration and fusion, image mapping, DEM generation and evaluation, digital mapping, data visualization and multimedia cartography, and electronic data access and dissemination.

Dr. Armenakis served as President of the Technical Commission IV on Spatial Information Systems and Digital Mapping of the International Society for Photogrammetry and Remote Sensing from 2000 to 2004. He is a member of the Canadian Institute of Geomatics and the American Society for Photogrammetry and Remote Sensing.

Deren Li received the B.Sc. and M.Sc. degrees from Wuhan University, Wuhan, China, and the Ph.D. degree from Stuttgart University of Germany, Stuggart, in 1963, 1981, and 1985, respectively, all in photogrammetry and remote sensing.

He is currently a Professor and Director of the National Key Laboratory for Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University. He has more than 350 technical papers, and eight books have been published on photogrammetry and remote sensing, image analyis, and geomatics.

Dr. Li has served as President of the Technical Commission III (1988–1992) and Commission VI (1992–1996) of International Society for Photogrammetry and Remote Sensing (ISPRS). He is an Academician of the Chinese Academy of Sciences and the Chinese Academy of Engineering, Academician of International Academy of European and Asian Studies.

Zhijun Wang received the B.Sc. degree in photogrammetry and remote sensing from the Wuhan University, Wuhan, China, the M.Sc. degree in geomatics from the International Institute for Geo-Information Science and Earth Observation, Enschede, The Netherlands, and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, China, in 1989, 1995, and 2001, respectively.

He is currently a Research Scientist of computer science at the University of Sherbrooke, Sherbrooke, CQ, Canada. His research interests include digital image processing, photogrammetry and remote sensing, geomatics, computer vision, and wavelet theory.

Dr. Wang holds an AUTO 21 Networks of Centres of Excellence Research Fellowship in the Image Analysis Laboratory, University of Sherbrooke.

Qingquan Li received the B.Sc., M.Sc., and Ph.D. degree in geomatics from Wuhan University, Wuhan, China, in 1985, 1988, 1998, respectively.

He is currently a Professor and Vice-President of Wuhan University. His research interests include digital image processing, photogrammetry and remote sensing, three-dimensional GIS, and LIDAR processing.