

A Model-Based Approach to Multiresolution Fusion in Remotely Sensed Images

Manjunath V. Joshi, *Member, IEEE*, Lorenzo Bruzzone, *Senior Member, IEEE*, and Subhasis Chaudhuri, *Senior Member, IEEE*

Abstract—In this paper, a model-based approach to multiresolution fusion of remotely sensed images is presented. Given a high spatial resolution panchromatic (Pan) image and a low spatial resolution multispectral (MS) image acquired on the same geographical area, the presented method aims to enhance the spatial resolution of the MS image to the resolution of the Pan observation. The proposed fusion technique utilizes the spatial correlation of each of the high-resolution MS channels by using an autoregressive (AR) model, whose parameters are learnt from the analysis of the Pan data. Under the assumption that the parameters of the AR model for the Pan image are the same as those that represent the MS images due to spectral correlation, the proposed technique exploits the learnt parameter values in the context of a proper regularization technique to estimate the high spatial resolution fields for the MS bands. This results in a combination of the spectral characteristics of the low-resolution MS data with the high spatial resolution of the Pan image. The main advantages of the proposed technique are: 1) unlike standard methods proposed in the literature, it requires no registration between the Pan and the MS images; 2) it models effectively the texture of the scene during the fusion process; 3) it shows very small spectral distortion (as it is less affected, compared to standard methods, by the specific digital numbers of pixels in the Pan image, since it exploits the learnt parameters from the Pan image rather than the actual Pan digital numbers for fusion); and 4) it can be used in critical situations in which the Pan and the MS images are acquired (also by different sensors) in slightly different areas. Quantitative experimental results obtained using Landsat-7 Enhanced Thematic Mapper Plus (ETM+) and Quickbird images point out the effectiveness of the proposed method.

Index Terms—Autoregressive model, combination, decimation, image processing, merging, model-based fusion, multiresolution fusion, multispectral image, panchromatic image, remote sensing, superresolution.

I. INTRODUCTION

THE PROCESS of combining panchromatic (Pan) and multispectral (MS) data to produce images characterized by both high spatial and spectral resolutions is known as multires-

olution fusion (see [1] for a precise definition of the problem, where it has been suggested to use the terms merging and combination in a much broader sense than fusion, with combination being even broader than merging). Given technological limitations related to the radiometric resolution of the detectors, MS images are generally acquired with a larger instantaneous field of view (IFOV) (i.e., lower spatial resolution) than Pan images. This depends on the relatively high spectral resolution of the MS channels, which should be balanced with a low spatial resolution to obtain an acceptable signal-to-noise ratio (SNR) in the acquired data. Due not only to this physical constraint, which involves a tradeoff between the spatial and the spectral resolutions, but also to the ever increasing availability of spaceborne sensors imaging with different scales and spectral bands, in the past few years, the remote-sensing community has devoted great attention to the process of multiresolution fusion. With the fusion of different images, we can overcome the limitations of information obtained from individual sources and obtain a better understanding of the observed scene [2]–[6].

Many researchers have addressed the problem of multiresolution image fusion for remote-sensing applications. The most common method for fusing a high-resolution Pan image with an MS remote-sensing image is based on the intensity–hue–saturation transform (IHS) technique [7]–[10]. This technique aims at enhancing the spatial details, but it also produces spectral distortion. High-pass filtering (HPF) fusion methods [11], [12], which are based on injecting high-frequency components into interpolated versions of MS data, have shown better performance by improving the quality of spatial details, resulting in a better visual effect. Other methods considered in the remote-sensing literature include approaches based on principal component analysis [13], [14] and the wavelet transform (WT) [15]–[18]. The standard WT method replaces the high-frequency component of the MS images with that of the Pan image by working in the wavelet domain and then synthesizing the fused image by applying the inverse WT. Fusion schemes based on the “à trous” wavelet (ATW) algorithm have also been proposed [19], [20]. In [21], Aiazzi *et al.* proposed two context-driven fusion methodologies based on the undecimated discrete WT (UDWT) and the generalized Laplacian pyramid (GLP), respectively. The former uses an octave bandpass representation obtained from a conventional WT by omitting the decimators and upsampling the wavelet filter bank. The latter is based on an oversampled structure obtained by recursively subtracting from an image an expanded and decimated low-pass version of it. Modeling the relationships

Manuscript received December 14, 2004; revised January 4, 2006. This work has been carried out under the framework of the India–Trento Program for Advanced Research.

M. V. Joshi is with the Dhirubhai Ambani Institute of Information and Communication Technology, Gujarat 382007, India (e-mail: mv_joshi@da-iict.org).

L. Bruzzone is with the Department of Information and Communication Technologies, University of Trento, 38050 Trento, Italy (e-mail: lorenzo.bruzzone@ing.unitn.it).

S. Chaudhuri is with the Department of Electrical Engineering, Indian Institute of Technology—Bombay, Mumbai 400076, India (e-mail: sc@ee.iitb.ac.in).

Digital Object Identifier 10.1109/TGRS.2006.873340

between wavelet coefficients of MS and Pan images at a coarser resolution and using them after extrapolation at a higher resolution for injecting high-frequency details into the MS images has been considered in [22], where an undecimated multiresolution approach to MS band sharpening is used. A common drawback of all the above methods is that they require accurate coregistration before the fusion process can take place. This may be a practical constraint affecting the fusion results, especially when considering Pan and MS images acquired by different satellites.

In this paper, we address the problem of multiresolution fusion from a different perspective, based on superresolution techniques [23], [24]. The idea of superresolution refers to the process of producing a high spatial resolution image from several low-resolution images, by increasing the maximum spatial frequency and removing the degradations that arise during image acquisition from a low-resolution camera. The most obvious superresolution method is to capture multiple low-resolution observations of the same scene by subpixel shift in the camera motion, i.e., using motion cue. However, this method requires an accurate registration process, which is a very difficult and challenging task. In order to overcome this difficulty, in recent years, image-processing researchers have started to exploit learning-based methods for image superresolution. In these methods, a database of high-resolution training images is used to create high-frequency details in zoomed images [25]–[27]. The advantage of learning-based methods is that they provide a very natural way of obtaining the required image characteristics. By choosing a proper feature set from training images, the quality of the obtained results can be improved. Nevertheless, instead of using a set of training images to learn the properties of the scenes, one could use available high-resolution observations to infer the required parameter values and exploit them along with low-resolution observations to accomplish the desired task. In [28] and [29], the authors use a multiresolution autoregressive (AR) model for the image, and exploit the extracted parameters for texture classification and image segmentation.

In this paper, we present a novel approach to multiresolution fusion, which exploits the available high-resolution Pan image to learn the spatial relationships for the unknown high-resolution MS images by using an AR structure. The AR parameters learnt from the Pan image are used as the AR parameters for the fused MS channels in order to enhance their spatial resolution. The assumption is that the same parameters learnt on the Pan image are valid for all MS channels, as we are only learning the spatial dependence between pixels and not their explicit values. This approach has four main advantages compared to classical multiresolution methods.

- 1) Unlike standard methods, it is totally insensitive to registration errors between the Pan and the MS images (this property is very important because a residual misregistration can be present between these images).
- 2) It shows very small spectral distortion (as it is less affected, compared to standard methods, by the specific digital numbers of pixels in the Pan image, since it exploits the learnt parameters from the Pan image rather than the actual Pan digital numbers for fusion).

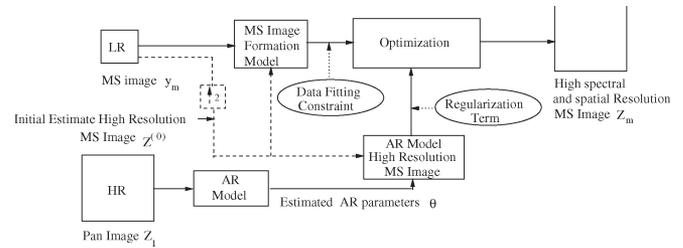


Fig. 1. Illustration of the data fusion for an MS and a Pan image for a 1 : 2 scale ratio. Here, \uparrow indicates upsampling, and LR and HR stand for low resolution and high resolution, respectively.

- 3) It models the texture of the analyzed scene effectively.
- 4) It can be applied to Pan and MS images acquired (also by different sensors) in slightly different geographical areas.

The outline of the paper is as follows. In Section II, a general description of the proposed method is presented. Section III-A discusses the modeling of the low spatial resolution MS image formation, while Section III-B addresses the AR model for the high-resolution image field. The estimation of model parameters used in the AR model of fused images is presented in Section IV-A. The regularization-based approach to enhance the spatial resolution of MS images is discussed in Section IV-B. Experimental results are given and analyzed in Section V. Finally, Section VI draws the conclusions of this paper.

II. GENERAL DESCRIPTION OF THE METHOD

The proposed method for multiresolution fusion is illustrated by the block scheme shown in Fig. 1. The low-resolution MS image is modeled by a simple decimation process. The decimation process relates to the fact that the low-resolution data are due to the integration of light falling on the charge-coupled device (CCD) sensor array of suitable area compared to the desired high-resolution images. We obtain the initial estimate of the high-resolution MS image by interpolating its low-resolution version using a proper interpolation technique. The contextual dependences in the unknown high-resolution MS image are modeled using an AR model. The required AR parameters are then estimated from the available high spatial resolution Pan image, and are used as coefficients for linear dependences in the AR model of the fused MS bands. This corresponds to injecting in the MS image the geometrical properties learnt from the high-resolution Pan observations. The basic assumption is that due to the high spatial correlation of each MS band with Pan data, the same parameterization can be used in all spectral channels. In greater detail, we define a cost function that consists of both a data-fitting term (based on the model of the low-resolution image and representing a constraint) and a regularization term (based on the AR model for the unknown high-resolution image). It is worth noting that the learnt AR parameters are used to enforce the spatial dependence in the regularization term. A suitable optimization technique is then exploited to minimize the cost function. The minimization leads to a high spatial resolution MS image that preserves the original spectral properties. The procedure is repeated for each MS channel using the same parameters estimated on the Pan data.

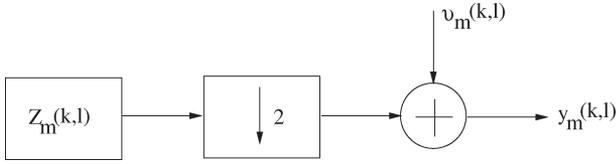


Fig. 2. Low-resolution image formation model illustrated for one of the spectral channels of an MS image.

Nevertheless, the effect of these parameters on different spectral bands may be significantly different because they represent only one term of the cost function to be minimized. In the following sections, the different parts of the proposed approach are described in greater detail.

It is worth noting that the proposed model-based approach differs from the *amelioration de la resolution spatiale par injection de structures* (ARSIS) model-based approach presented in [31], where the authors use a wavelet-based multiresolution analysis. They assume a suitable model to infer the unknown wavelet coefficients for the MS images, taking into account the physics of Pan and MS images and the correlation between their computed wavelet coefficients at lower scales. These wavelet coefficients are used to derive the model parameters that are then used for obtaining the missing wavelet coefficients of the MS images at the higher scale. On the contrary, the proposed approach uses the AR model for the given high spatial resolution Pan image as well for the unknown high-resolution MS images. Then, the model parameters learnt from the Pan image are used in a suitable regularization framework to obtain high spatial and high spectral resolution MS images without requiring any registration between MS and Pan images.

III. MS AND PAN IMAGE MODELS

A. MS Image Formation Model

The multiresolution fusion problem can be cast in a high-resolution restoration framework, where an appropriate prior model learnt using the high spatial resolution Pan observations is exploited to regularize the solution. Let us consider p images, made up of a single high-resolution Pan image and $p - 1$ low-resolution MS channels. Fig. 2 shows the block scheme of the model that represents the relationship between the available low-resolution MS image and its high-resolution version for a single spectral channel. In particular, the observed low-resolution MS images are modeled as decimated and noisy versions of their high-resolution versions (to be estimated). Let \mathbf{z}_1 represent the available lexicographically ordered high-resolution Pan image.

If $\mathbf{z}_m, m = 2, 3, \dots, p$, represents the lexicographically ordered high spatial resolution MS images (each of $N^2 \times 1$ pixels) and $\mathbf{y}_m, m = 2, 3, \dots, p$, are the corresponding vectors (of size $M^2 \times 1$) containing pixels from the low spatial resolution observations, then we can write

$$\mathbf{y}_m = D\mathbf{z}_m + \nu_m, \quad m = 2, 3, \dots, p \quad (1)$$

where D is the decimation matrix of size $M^2 \times N^2$ and ν_m is the $M^2 \times 1$ noise vector that we assume to be a zero-mean independent and identically distributed (i.i.d.) process

with variance σ^2 . For a decimation (downsampling) factor of q , the decimation matrix D consists of q^2 nonzero elements of value $1/q^2$ along each row at appropriate locations [32]. As an example, for a decimation factor of $q = 2$ and with lexicographically ordered \mathbf{z}_m of size 16×1 , the D matrix is of size 4×16 and can be written as

$$D = \frac{1}{4} \begin{bmatrix} 1100110000000000 \\ 0011001100000000 \\ 0000000011001100 \\ 0000000000110011 \end{bmatrix}. \quad (2)$$

Thus, the low-resolution intensity is the average of the high-resolution intensities over a neighborhood of q^2 pixels corrupted with additive noise. This decimation model simulates the integration of light intensity that falls on the high-resolution detector, and the decimation process is represented by the matrix D , which has the structure given in (2).

Our problem now reduces to estimating $\mathbf{z}_m, m = 2, \dots, p$, given $\mathbf{y}_m, m = 2, \dots, p$, and \mathbf{z}_1 (high-resolution Pan image). This is an ill-posed inverse problem. The problem is ill posed as the matrix D is not invertible, i.e., one is required to estimate values of four pixels from a single pixel when the upsampling factor $q = 2$. It should be mentioned here that the aliasing process modeled by (1) accounts for the spatial frequency distortion, conserving the spectral properties of the MS images to be fused. This is a very important feature of the proposed technique.

B. High-Resolution Image Field Modeling

As discussed in Section III-A, we consider this as an ill-posed image-fusion problem as there are infinite solutions to (1). Thus, obtaining a robust solution requires a reasonable assumption (constraint) about the nature of the true image. Once the prior model for the true image is chosen, the obtained solution depends on the model parameters. A proper choice of the model parameters leads to a better solution and ill-posedness results in a better posed problem. However, the parameters of the prior model are unknown as the true high-resolution MS images are unavailable and have to be estimated. To solve this ambiguity, we propose to use the available high-resolution Pan observations of the same scene for the learning of the parameter values, so that they can be used to improve the solution. In other words, the spatial correlation for the unknown high-resolution MS spectral channels can be estimated from the available high-resolution Pan data. In order to accomplish this task, a proper model for the high-resolution Pan image is needed from which to extract parameters. These parameters characterize the mutual influence among context-dependent entities, such as pixel intensities and other spatially correlated features. A possible approach to model spatial dependence is to use Markov random fields (MRFs) [33], [34]. MRFs achieve this goal by characterizing mutual influence among the entities using conditional probabilities for a given neighborhood. However, although the MRF model for priors is a popular statistical model (and effectively captures contextual dependences), it involves very high computational

| | | | | |
|---------|---------|---------|---------|---------|
| (-2,-2) | (-2,-1) | (-2, 0) | (-2, 1) | (-2, 2) |
| (-1,-2) | (-1,-1) | (-1, 0) | (-1, 1) | (-1, 2) |
| (0, -2) | (0, -1) | (0, 0) | (0, 1) | (0, 2) |
| (1, -2) | (1, -1) | (1, 0) | (1, 1) | (1, 2) |
| (2, -2) | (2, -1) | (2, 0) | (2, 1) | (2, 2) |

Fig. 3. Fifth-order neighborhood for the pixel at location (0, 0).

complexity, as one needs to compute the partition function in order to estimate the true parameters. The computational burden can be reduced if we use a scheme such as the maximum pseudolikelihood (MPL), which does not require the partition functions to be estimated [35]. Nonetheless, to obtain the global minimum, we still need to use a stochastic relaxation technique, which is computationally taxing. In addition, pseudolikelihood is not a true likelihood except for the trivial case of a null neighborhood. This motivates us to use a different (but proper) prior. We assume a linear dependence of a pixel to its neighbors in a high spatial resolution image and represent it with an AR model. Although this becomes a weaker model than the MRF-based one, it is more suitable to process the large amount of data involved in remote-sensing applications as the computation is drastically reduced. It is worth noting that an MRF model can represent a large variety of spatial interrelationships among pixels locally, out of which the AR model is only a subset representing linear dependences. Interested readers are referred to [36] for details.

Let $z(s)$ be the gray-level value of the image pixel at site $s = (i, j)$ in an $N \times N$ lattice, where $i = 1, 2, \dots, N$ and $j = 1, 2, \dots, N$. The AR model for $z(s)$ can be expressed as [29]

$$z(s) = \sum_{r \in \mathcal{N}_s} \theta(r)z(s+r) + \sqrt{\rho}n(s) \quad (3)$$

where \mathcal{N}_s is the neighborhood of pixel at s . $\theta(r)$, r being a neighborhood index with $r \in \mathcal{N}_s$, and ρ are unknown parameters, and $n(\cdot)$ is an i.i.d. noise sequence with zero mean and unit variance. Here, ρ is the variance of the white noise that generates the specified data for the given AR parameters. We use a fifth-order neighborhood that requires a total of 24 parameters $\theta(i, j)$ (see Fig. 3). The choice of neighborhood is motivated by the fact that, on the one hand, a larger neighborhood captures the local texture very well but then it becomes too specific and lacks the generalization capability. If the local texture somehow differs from the globally estimated texture, the model fits very poorly at those sites. On the other hand, a small size of neighborhood generalizes very well, but lacks the capability to capture the intricacies of the image texture. Thus, a neighborhood size of five appears to represent a good trade-off. Furthermore, it is commonly observed in many satellite images that the scene content does not have direction-specific textures. This means that there is no left or right or up or down preference while modeling the texture in a given image. Under such cases, one may assume the AR model to be rotationally symmetric, thus reducing the number of unknown parameters.

In view of above and to reduce the computation time taken to estimate these parameters, we use a symmetric AR model, where $\theta(r) = \theta(-r)$, and estimate only eight parameters. We observed that the improvement in the performance is negligible when we considered an asymmetric AR model with all the 24 parameters estimated, but the computational complexity increases marginally. Nonetheless, we show results of fusion for both cases in the experimental section. It may be mentioned here that we assume a folded torus model of the image [29] so that border pixels are taken care of while estimating the parameters.

IV. PARAMETER ESTIMATION AND FUSION

A. Estimation of AR Prior Parameters

One of the characteristics of an image is the statistical dependence of the gray level at a lattice point on those of its neighbors. As stated above, we characterize this statistical dependence by using an AR model, where the gray level at a location is expressed as a linear combination of the neighborhood gray levels and an additive noise. We estimate the AR model parameters by considering the image as a finite lattice model using the iterative scheme given in [29]. In this way, we use the Pan image \mathbf{z}_1 as an AR model, considering the same neighborhood size around each pixel. Thus, we use a homogeneous model and derive a set of parameters for the entire Pan image. Once we estimate the values of the AR model parameters θ , we use them for the fusion of each of the MS channels. In other words, we extract parameters from the high-resolution Pan image and use the same AR prior for the unknown high-resolution MS images. The extracted AR parameters can also be used on MS spectral channels with bandwidths outside the Pan image (even though this is a suboptimal application of the proposed technique as, depending on the considered area, the texture properties of the spectral channels outside the range of wavelengths of the Pan image may be only partially correlated with the texture properties of the Pan data). Another attractive scenario is the one in which the fusion task is extended to portions of the MS image for which the Pan observations are not available. A critical issue may be the stationarity of the model parameters versus the portion of the considered Pan area. To solve this problem, the approach could be applied to subimages with similar properties where we can reasonably assume spatial stationarity.

B. Fusion Using AR Prior

Regularization is a popular method for interpolating sparse data, as well as smoothing the data obtained from noisy measurements. The regularization-based approach is quite amenable to the integration of information from multiple observations with the smoothness function chosen from the prior knowledge of the function to be estimated here. The AR prior knowledge plays the role of a contextual constraint used to regularize the solution. The AR parameters estimated from the Pan observations are used in the cost function that is to be minimized. The minimization leads to estimates of

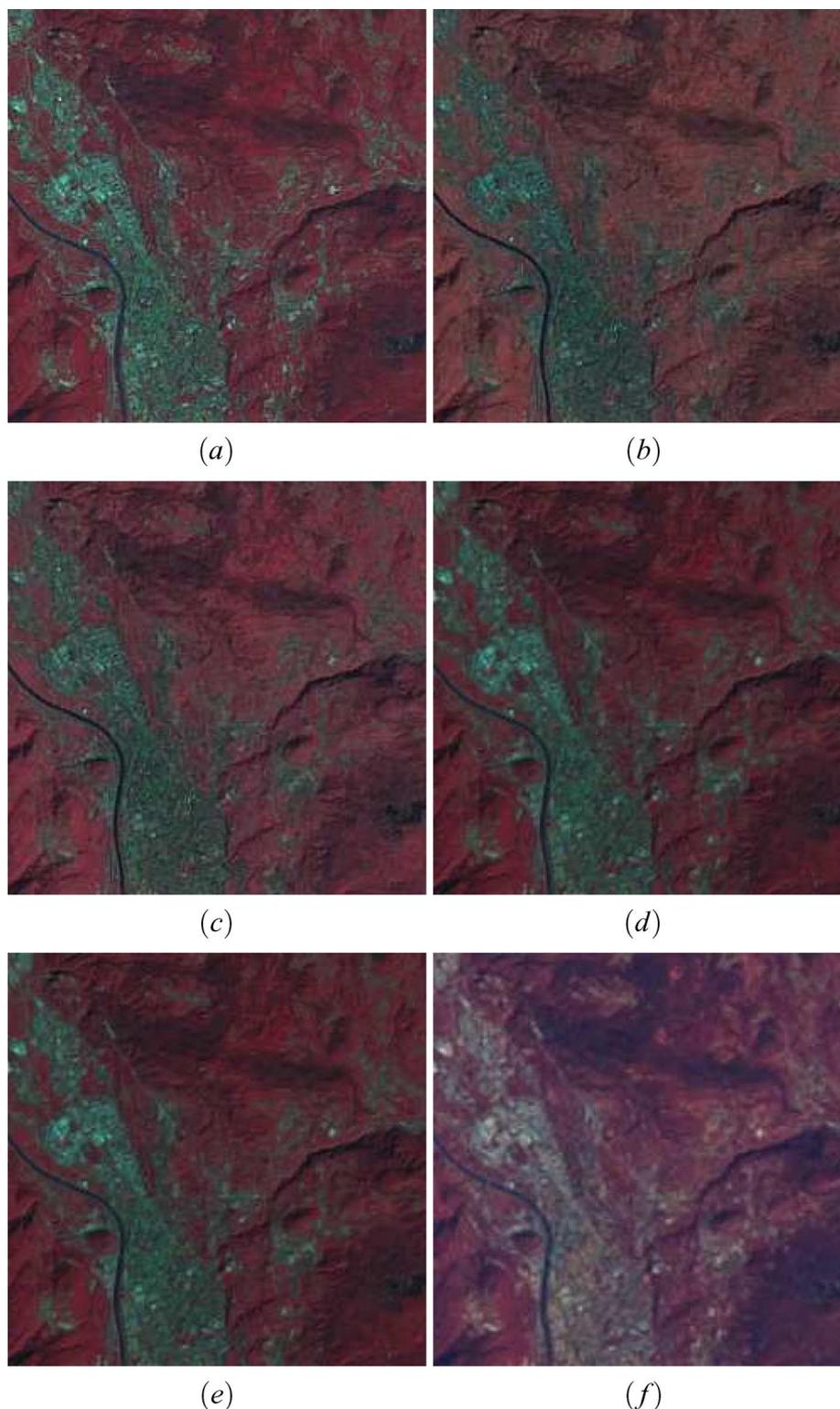


Fig. 4. RGB false-color composition 4, 3, and 2 of fusion results at $30\text{ m} \times 30\text{ m}$ obtained from degraded ETM+ MS bands (Landsat-7 data set). (a) True MS image. (b) HPF fusion. (c) IHS fusion. (d) BT fusion. (e) GS method. (f) Bicubically interpolated MS image.

high spatial resolution MS images. We use the simple linear dependence of a pixel value on its neighbors as a constraint using the AR model for the high spatial resolution MS image to be recovered. The AR set of parameters θ estimated from the Pan image contains the coefficients for linear dependence in the prior term. Using a data-fitting term [see (1)] and a prior term

[see (3)], one can easily derive the corresponding cost function to be minimized as

$$\varepsilon_m = \lambda \|y_m - D\mathbf{z}_m\|^2 + \sum_{i,j} \left(z_m(s) - \sum_{r \in \mathcal{N}_s} \theta(r) z_m(s+r) \right)^2 \tag{4}$$

where $m = 2, 3, \dots, p$. Here, λ is a regularization parameter that is proportional to σ^2/ρ , where ρ is the variance of the driving noise source in (3) and σ^2 is the variance at the image sensor, as given in (1) (i.e., the data-fitting error). Note that the first term in the above equation relates to the confidence in data fitting, and the second term relates to the confidence in model fitting. The second term provides the spatial regularization while upsampling the observations. Since the model parameter vector θ has already been estimated from the Pan image, as discussed in the previous section, a solution to the above equation can be obtained. It is worth noting that the above cost function is convex and therefore can be minimized using the gradient-descent technique. In order to speed up the computation, the initial estimate $\mathbf{z}_m^{(0)}$ is obtained as the bicubically interpolated version of the MS image \mathbf{y}_m . The estimation process converges in a few iterations, making the procedure computationally very efficient. The minimization is carried out independently for each of the MS bands using the same parameter set learnt from the Pan data, resulting in the fused high spatial and high spectral resolution MS images.

V. EXPERIMENTAL RESULTS AND COMPARISONS

In this section, to assess the effectiveness of the proposed approach, we consider two different data sets acquired by Landsat-7 Enhanced Thematic Mapper Plus (ETM+) and QuickBird satellite sensors, respectively. The purpose of selecting such different types of sensors is to demonstrate that the proposed method works for both moderately high and very high spatial resolution data. All experiments were carried out on images acquired over the same geographical area at the same time. The trials have been conducted on spatially degraded images so that the effectiveness of the proposed technique can be assessed quantitatively by analyzing the fused MS products and the true MS images. A set of quality assessment measures, which are currently used in the literature, has been selected for quantitative analysis. It is worth noting that, in this paper, we do not investigate the critical issue of quality assessment of fused images. Interested readers are referred to [22].

A. Results on Landsat-7 ETM+ Images

First of all, we consider the fusion of a Landsat-7 ETM+ MS image with the corresponding coregistered Pan image. For this experiment, we used a high spatial resolution Pan image and six spectral channels acquired over the city of Trento, Italy. The MS image include the visible, near-IR (NIR), and mid-IR bands, while the Pan image covers the spectral range given by 520–900 nm. The data are resampled to a ground resolution of $30 \text{ m} \times 30 \text{ m}$ and $15 \text{ m} \times 15 \text{ m}$ ground sample distance (GSD) for MS and Pan, respectively, yielding a decimation factor $q = 2$ between them. They are then degraded in order to obtain $60 \text{ m} \times 60 \text{ m}$ and $30 \text{ m} \times 30 \text{ m}$, MS and Pan images, respectively. These degraded images are then used for fusion, using the proposed approach to estimate the $30 \text{ m} \times 30 \text{ m}$ resolution fused MS images. The true Pan and the MS images are of size 512×512 and 256×256 pixels, respectively, and the corresponding degraded versions used in



Fig. 5. RGB false-color composition 4, 3, and 2 of fusion results at $30 \text{ m} \times 30 \text{ m}$ obtained from degraded ETM+ MS bands with the proposed model-based approach (Landsat-7 data set).

the experiments are of size 256×256 and 128×128 pixels, giving a decimation factor of two. The proposed method is compared with the HPF, the IHS, the Gram–Schmidt (GS), and the Brovey transform (BT) techniques. A comparison is also shown with the bicubically interpolated MS images. A 3×3 window size filter is exploited for the low-pass filtering operation in the HPF technique. Concerning the GS method, we adopted the algorithm implemented in the commercial Envi tool [37]. The required input for the low-resolution Pan image was a decimated version of the true Pan image. For the proposed method, the AR parameters from the Pan image were estimated and used in the cost-function minimization. The more proper value of ρ in (3) for this experiment was equal to 0.0063. The value of the regularization parameter λ , which tunes the “smoothness” of the solution, was set to $\lambda = 0.5$ through a trial-and-error procedure. It is worth noting that one could use a generalized cross-validation technique to identify the best value of λ , but this is computationally very demanding and is specific to a given image only. A step size of 0.05 was used in the gradient-descent algorithm, and the maximum number of iterations was set to 100. However, the optimization process converged much before the set limit. The processing time on a Pentium IV 2.0-GHz machine was only a few seconds, a little higher than for conventional fusion technique such as the HPF-based method. However, our comparison does not include the computational time required for coregistration of different bands, which is a must for other literature techniques. For these reasons, the proposed method compares very favorably with other standard methods in terms of computational requirements. The performance assessment task is accomplished according to both a subjective visual inspection and a quantitative analysis of the statistical properties of the fused images.

In order to assess the effectiveness of the proposed technique, in the first set of experiments, we considered the fusion of visible and NIR (VNIR) MS bands, whose total spectral range is the same as that of the Pan image. Fig. 4(a) shows the RGB false-color composition of VNIR spectral bands 4, 3, and 2 of the original image at $30 \text{ m} \times 30 \text{ m}$ resolution. From the spectral characteristics of the ETM+ images [38], one can see that these

TABLE I

QUANTITATIVE ASSESSMENT OF THE FUSION RESULTS PROVIDED BY THE CONSIDERED TECHNIQUES APPLIED TO THE LANDSAT-7 ETM+ DATA SET: (a) BAND 2, (b) BAND 3, (c) BAND 4 (THE IHS AND BT TECHNIQUES HAVE BEEN APPLIED TO BANDS 2, 3, AND 4 FALSE-COLOR COMPOSITIONS, WHEREAS THE OTHER TECHNIQUES HAVE BEEN APPLIED TO EACH SPECTRAL CHANNEL SEPARATELY)

| Image | Std Dev | Definition | Distortion extent | Deviation index | Corr coeff | MSE |
|-----------|---------|------------|-------------------|-----------------|------------|-------|
| MS-Band 2 | 13.204 | 3.426 | - | - | - | - |
| Bicubic | 12.055 | 1.769 | 2.956 | 0.053 | 0.939 | 0.008 |
| HPF | 12.822 | 3.424 | 4.272 | 0.080 | 0.895 | 0.013 |
| IHS | 9.241 | 3.270 | 7.275 | 0.131 | 0.697 | 0.040 |
| BT | 9.009 | 3.114 | 7.185 | 0.128 | 0.694 | 0.041 |
| GS | 12.307 | 2.773 | 3.896 | 0.072 | 0.899 | 0.012 |
| Proposed | 12.410 | 2.113 | 2.848 | 0.052 | 0.946 | 0.006 |

(a)

| Image | Std Dev | Definition | Distortion extent | Deviation index | Corr coeff | MSE |
|-----------|---------|------------|-------------------|-----------------|------------|-------|
| MS-Band 3 | 19.297 | 4.878 | - | - | - | - |
| Bicubic | 17.694 | 2.549 | 4.136 | 0.082 | 0.943 | 0.017 |
| HPF | 18.266 | 3.979 | 5.230 | 0.111 | 0.922 | 0.024 |
| IHS | 11.646 | 3.455 | 9.325 | 0.198 | 0.760 | 0.069 |
| BT | 12.823 | 3.186 | 7.490 | 0.141 | 0.853 | 0.058 |
| GS | 17.856 | 3.678 | 5.407 | 0.114 | 0.909 | 0.026 |
| Proposed | 18.192 | 3.022 | 3.893 | 0.079 | 0.952 | 0.014 |

(b)

| Image | Std Dev | Definition | Distortion extent | Deviation index | Corr coeff | MSE |
|-----------|---------|------------|-------------------|-----------------|------------|-------|
| MS-Band 4 | 15.831 | 4.705 | - | - | - | - |
| Bicubic | 14.497 | 2.605 | 4.016 | 0.064 | 0.944 | 0.006 |
| HPF | 15.291 | 3.987 | 5.040 | 0.080 | 0.912 | 0.009 |
| IHS | 18.779 | 4.261 | 7.857 | 0.127 | 0.861 | 0.024 |
| BT | 18.827 | 4.672 | 8.371 | 0.132 | 0.848 | 0.025 |
| GS | 15.280 | 4.475 | 5.616 | 0.090 | 0.884 | 0.011 |
| Proposed | 14.982 | 3.099 | 3.710 | 0.059 | 0.953 | 0.005 |

(c)

bands overlap very well with the Pan band, justifying the use of AR parameters from the Pan image. As can be observed from the analysis of the figure, the fused image obtained with the proposed model-based approach (see Fig. 5) is very similar to the original one, without any significant spectral distortion. In particular, one can see that the degree of detail in the urban area and in the roads is sharply increased compared to the original degraded image [for reference, see the bicubically interpolated MS image in Fig. 4(f)]. The IHS and BT fusion images shown in Fig. 4(c) and (d) lack the spatial details and also reveal significant spectral distortions. The results provided by the HPF fusion technique [see Fig. 4(b)] are comparable to those yielded by the proposed approach as far as spatial details are concerned. In greater detail, the HPF method enhances the edges, resulting in fused images that look sharper due to HPF on the Pan image. However, this also increases the noise. On the contrary, the proposed method does not enhance the noise; rather, it works like a bandpass filter that boosts up the high-frequency content without increasing the noisy components. Concerning the GS

TABLE II

GLOBAL QUANTITATIVE ASSESSMENT OF THE FUSION RESULTS IN TERMS OF RASE AND ERGAS FOR BANDS 4, 3, AND 2 OF THE LANDSAT ETM+ DATA SET

| Image | RASE | ERGAS |
|----------|-------|-------|
| Bicubic | 0.645 | 3.178 |
| HPF | 0.645 | 3.522 |
| BT | 0.729 | 4.418 |
| IHS | 0.724 | 4.499 |
| GS | 0.709 | 4.089 |
| Proposed | 0.593 | 2.473 |

method, it slightly improved detail modeling with respect to the proposed method, but significantly increased spectral distortions. It is worth noting that in our approach, we apply the fusion algorithm to each spectral channel separately and not to the RGB color composition. We use the RGB composition only to visualize the results and compare different fusion methods.

An objective quantitative assessment of the quality of MS images sharpened using Pan data is a difficult task. Even when reference images are available for comparison, fidelity to the reference usually requires the computation of a number of statistical parameters. According to the existing literature, in order to quantify performances obtained on each spectral channel, we use the following set of statistical parameters: 1) standard deviation; 2) definition of the original image and the fused images; 3) distortion extent; 4) deviation index; and 5) correlation coefficient between the fused and the MS images. The mathematical descriptions of these performance indexes can be found in [39]. The comparison is also made on the basis of the mean squared error (mse) between the true MS images and the fused images. The mse is defined as

$$mse = \frac{\sum_i (f_i - \hat{f}_i)^2}{\sum_i f_i^2} \tag{5}$$

where f_i represents the true MS image intensity value at the i th pixel and \hat{f}_i is the corresponding fused MS image intensity. The comparison is done considering each band separately. We also evaluate the global spectral fidelity of the merged images computed on different spectral channels in terms of the *relative average spectral error* (RASE) index, expressed as a percentage, as well as the *erreur relative globale adimensionnelle de synthèse* (ERGAS) index [40].

The standard deviation and definition parameters indicate the evidence of quality in terms of spatial details, while the distortion extent, deviation index, and correlation values reflect the fidelity of spectral information. For example, the mse, distortion, and the deviation index should be as low as possible. On the contrary, the higher the definition or correlation coefficient, the better is the fusion. The obtained results are shown in Tables I and II. As can be observed from the tables, the proposed approach shows better performances in terms of spectral fidelity, by significantly reducing spectral distortion compared to other methods. It is worth noting that the model-based technique shows the maximum value of correlation, and the minimum distortion and deviation values, as well as the

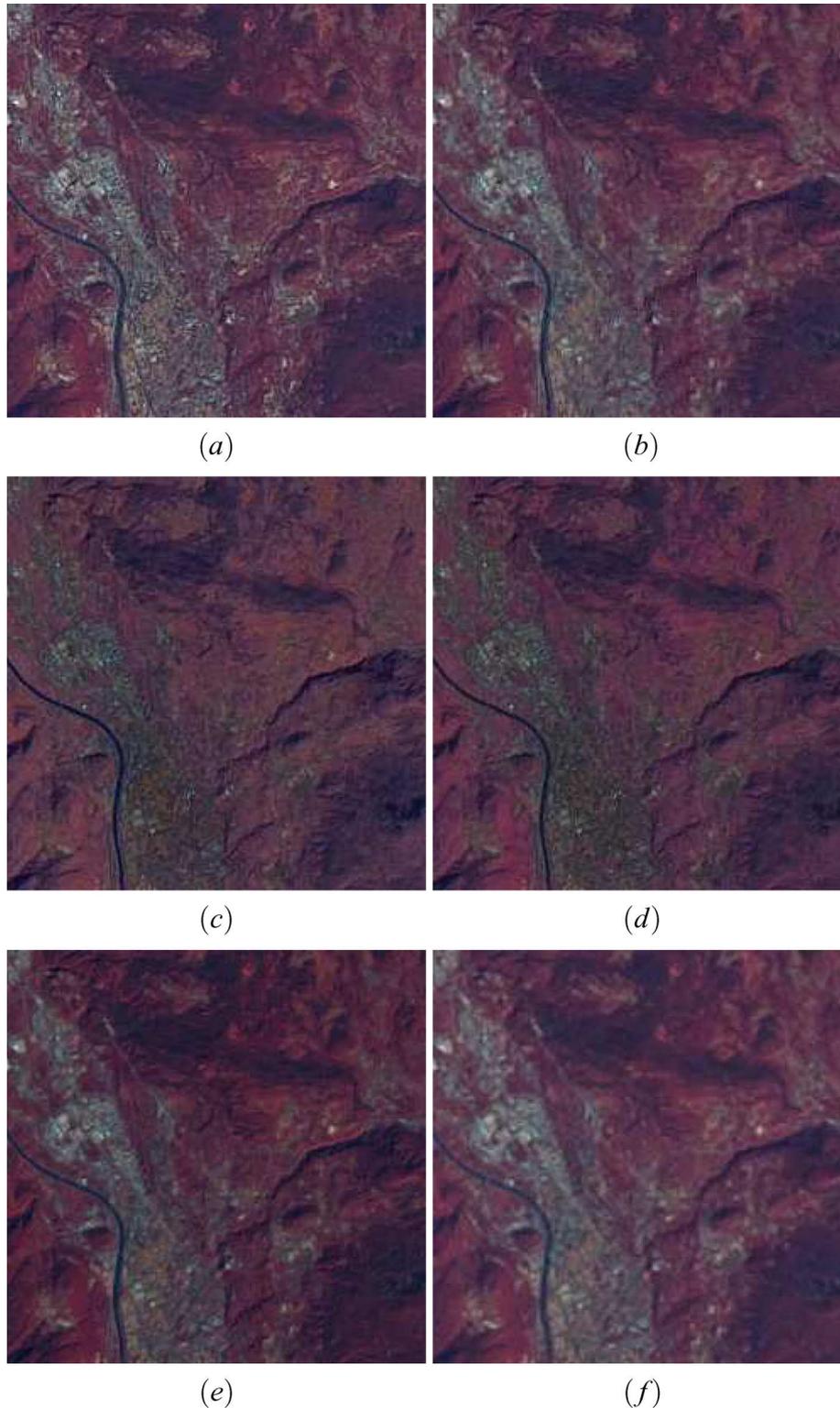


Fig. 6. RGB false-color composition 5, 3, and 1 of fusion results at $30\text{ m} \times 30\text{ m}$ obtained from degraded ETM+ MS bands (Landsat-7 data set). (a) True MS images. (b) HPF fusion. (c) IHS fusion. (d) BT fusion. (e) GS method. (f) Bicubically interpolated MS image.

lowest values of the RASE and ERGAS indexes. In addition, it results in a substantial improvement in the mse compared to other methods. Finally, the standard deviation and definition parameter values, even if not the best ones among those provided by the considered techniques, are also reasonably close to the true MS image statistics.

In the second set of experiments, we repeated the previous trials in the case where the spectral range of the considered MS bands was different from that of the Pan image. In particular, we considered the Pan sharpening of ETM+ spectral bands 5, 3, and 1. This test is particularly critical for the IHS technique, which explicitly assumes the use of MS bands in

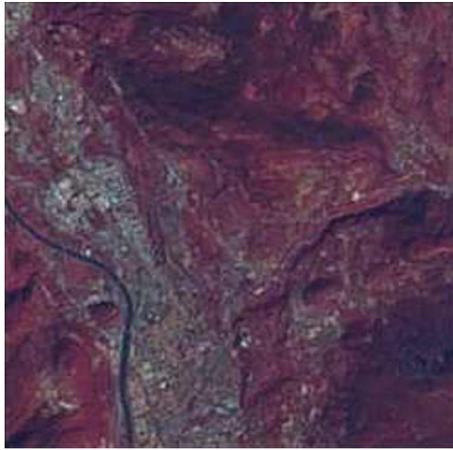


Fig. 7. RGB false-color composition 5, 3, and 1 of the fusion results at 30 m × 30 m obtained from degraded ETM+ MS bands with the proposed model-based approach (Landsat-7 data set).

the spectral range of Pan. However, band 5 (155–175 nm) is outside the spectral range of Pan [38]. Concerning the proposed technique, the AR parameters learnt from the Pan image were used again in the optimization technique to minimize the cost function. The various parameters used in the gradient-descent algorithm are kept the same as in the previous experiment. Fig. 6(a) shows the RGB false-color composition of original spectral channels 5, 3, and 1 at 30 m × 30 m resolution. Figs. 6(b)–(e) and Fig. 6(f) display the fusion results using the reference techniques considered and bicubically interpolated MS images, respectively. The results using the proposed approach are shown in Fig. 7. Quantitative measures are given in Tables III and IV. As can be seen from the analysis of these figures and tables, the proposed method provides good performances even when the estimated AR parameters derived from the Pan image are used for Pan sharpening of the MS bands that lie outside the spectral range of the Pan image. As an example, a quantitative comparison of results obtained by the proposed technique with those provided by the HPF method for band 1 [see Table III(a)] shows that the correlation coefficient and the deviation index for the proposed method are 0.941 and 0.039, respectively, while they are 0.886 and 0.061, respectively, for the HPF technique. In addition, the RASE and ERGAS values associated with the proposed approach are the lowest ones among the considered techniques. This confirms that the proposed method exhibits better capabilities in preserving spectral properties. As regards the performances in terms of standard deviation and definition (which are related to spatial fidelity), the GS and the HPF methods gave slightly better results (but a perfect alignment between Pan and MS images is required). It may be noted here that the statistical parameters for MS band 3 given in Tables I(b) and III(b) for different methods are the same, except for the IHS and the BT techniques, as these methods use color compositions 4, 3, and 2, and 5, 3, and 1, respectively, for fusion.

B. Results on Quickbird Images

In this section, we describe and analyze the results obtained on a Quickbird data set consisting of four MS bands at a spatial

TABLE III
QUANTITATIVE ASSESSMENT OF THE FUSION RESULTS PROVIDED BY THE CONSIDERED TECHNIQUES APPLIED TO THE LANDSAT ETM+ DATA SET: (a) BAND 1, (b) BAND 3, (c) BAND 5 (THE IHS AND BT TECHNIQUES HAVE BEEN APPLIED TO BANDS 1, 3, AND 5 FALSE-COLOR COMPOSITIONS, WHILE THE OTHER TECHNIQUES HAVE BEEN APPLIED TO EACH SPECTRAL CHANNEL SEPARATELY)

| Image | Std Dev | Definition | Distortion extent | Deviation index | Corr coeff | MSE |
|-----------|---------|------------|-------------------|-----------------|------------|-------|
| MS-Band 1 | 12.238 | 3.241 | - | - | - | - |
| Bicubic | 11.153 | 1.548 | 2.801 | 0.040 | 0.935 | 0.004 |
| HPF | 11.923 | 3.313 | 4.141 | 0.061 | 0.886 | 0.008 |
| IHS | 7.273 | 3.293 | 11.739 | 0.165 | 0.205 | 0.062 |
| BT | 7.895 | 3.555 | 12.508 | 0.177 | 0.171 | 0.068 |
| GS | 11.220 | 2.209 | 3.437 | 0.049 | 0.903 | 0.006 |
| Proposed | 11.328 | 1.855 | 2.757 | 0.039 | 0.941 | 0.004 |

(a)

| Image | Std Dev | Definition | Distortion extent | Deviation index | Corr coeff | MSE |
|-----------|---------|------------|-------------------|-----------------|------------|-------|
| MS-Band 3 | 19.297 | 4.878 | - | - | - | - |
| Bicubic | 17.694 | 2.549 | 4.136 | 0.082 | 0.943 | 0.017 |
| HPF | 18.266 | 3.979 | 5.230 | 0.111 | 0.922 | 0.024 |
| IHS | 10.413 | 3.383 | 11.178 | 0.218 | 0.643 | 0.110 |
| BT | 9.965 | 2.749 | 10.305 | 0.188 | 0.790 | 0.102 |
| GS | 17.856 | 3.678 | 5.407 | 0.114 | 0.909 | 0.026 |
| Proposed | 18.192 | 3.022 | 3.893 | 0.079 | 0.952 | 0.014 |

(b)

| Image | Std Dev | Definition | Distortion extent | Deviation index | Corr coeff | MSE |
|-----------|---------|------------|-------------------|-----------------|------------|-------|
| MS-Band 5 | 20.244 | 6.876 | - | - | - | - |
| Bicubic | 17.912 | 3.426 | 5.887 | 0.090 | 0.918 | 0.012 |
| HPF | 18.621 | 4.586 | 6.707 | 0.101 | 0.894 | 0.015 |
| IHS | 16.855 | 4.139 | 12.914 | 0.177 | 0.700 | 0.059 |
| BT | 16.448 | 4.334 | 14.107 | 0.190 | 0.656 | 0.068 |
| GS | 19.221 | 6.008 | 8.127 | 0.125 | 0.843 | 0.023 |
| Proposed | 18.698 | 4.177 | 5.481 | 0.082 | 0.930 | 0.010 |

(c)

TABLE IV
GLOBAL QUANTITATIVE ASSESSMENT OF THE FUSION RESULTS PROVIDED BY THE CONSIDERED TECHNIQUES IN TERMS OF RASE AND ERGAS MEASURES FOR BANDS 5, 3, AND 1 OF THE LANDSAT-7 ETM+ DATA SET

| Image | RASE | ERGAS |
|----------|-------|-------|
| Bicubic | 0.618 | 2.951 |
| HPF | 0.744 | 3.482 |
| BT | 1.282 | 5.712 |
| IHS | 1.229 | 5.624 |
| GS | 0.965 | 4.399 |
| Proposed | 0.425 | 2.076 |

resolution of 2.4 m × 2.4 m and a coregistered Pan image with a spatial resolution of 0.6 m × 0.6 m (giving a spatial resolution difference by a factor of four). The images have size 128 × 128 pixels and 512 × 512 pixels for MS bands and Pan, respectively, and were acquired at the same time over the Malpensa area, Italy. Since the MS bands cover a total spectral range of 450–900 nm, in this case, for fusion, we use the three

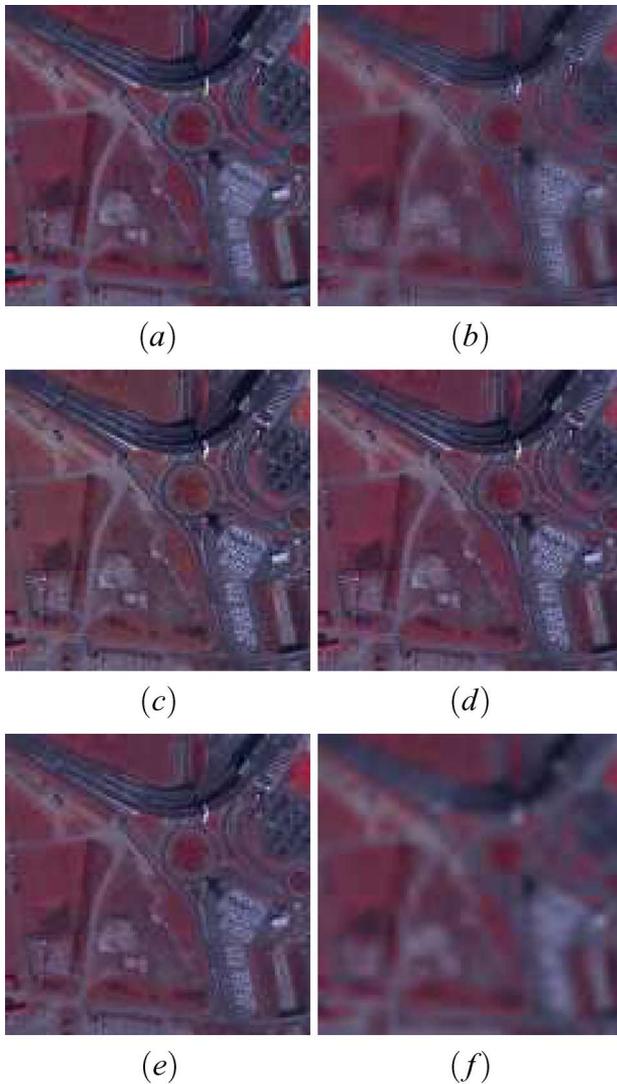


Fig. 8. Fusion results at $2.4 \text{ m} \times 2.4 \text{ m}$ obtained from the degraded bands 4, 3, and 2 images shown as a color composite (Quickbird data set). (a) Image formed from true bands 4, 3, and 2. (b) HPF fusion. (c) IHS fusion. (d) BT fusion. (e) GS method using the standard Envi package. (f) Bicubically interpolated image.

MS NIR, blue, and green bands, i.e., bands 4, 3, and 2, as they all are within the spectral range of the Pan image [41]. The Pan image decimated by a factor of four was used to learn the AR parameters, and the original MS images were used as reference (true) data for comparisons. In order to make a quantitative assessment of performances possible, as in the previous data set, all trials were carried out on degraded versions of the images (decimated by a factor of four).

We report the color composition of fused images obtained for the bands 4, 3, and 2 by standard fusion methods [Fig. 8(b)–(e)], bicubical interpolation [Fig. 8(f)], and the proposed model-based method [Fig. 9(a)]. Fig. 8(a) shows the true-color composite image for bands 4, 3, and 2 at $2.4 \text{ m} \times 2.4 \text{ m}$ resolution. Upon comparison of the figures, it can be seen that the details in the fused image obtained with the proposed approach are very similar to those in the true MS image, while in the other methods, the texture only resembles that of the Pan image. The reason is that all the other methods use the

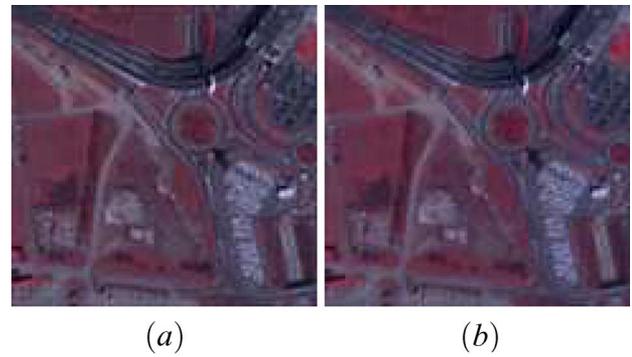


Fig. 9. Fusion result at $2.4 \text{ m} \times 2.4 \text{ m}$ obtained from the degraded MS bands 4, 3, and 2 images by the proposed model-based approach with. (a) Symmetric AR model (eight parameters). (b) Asymmetric model (24 parameters).

specific digital numbers (DNs) of the Pan image for fusion, while the proposed approach works on the principle of spatial relationship on a neighborhood by learning the AR parameters rather than using the DN of the Pan data. The quantitative measures given in Tables V(a)–(c) and VI confirm the superiority of the proposed approach in all cases as regards spectral fidelity [see the last four columns of Table V(a)–(c)]. Concerning the spatial fidelity, in the case of perfect geometrical alignment between the MS and Pan images, the statistics given in the first two columns of the table point out that the proposed approach provides either the best results or results quite close to the best, the best figure being obtained by different methods for different bands. These results are particularly promising for the proposed method, especially considering that it does not require an accurate spatial registration between the Pan and the MS images, and that the learning phase of the AR model parameters could also be carried out on a Pan image acquired in a slightly different area than that showed in the MS image.

In order to better understand the properties of the proposed approach, experiments were carried out to assess the robustness of the presented model-based method with respect to misregistration errors in very high geometrical resolution images. To this purpose, we simulated a misregistration of (2, 2) pixels for all the MS images and analyzed the performances provided by the presented model-based method with respect to other methods (see Table V). The results related to misregistered images are indicated with the term “misregistered” in the tables. It is worth noting that the fused images obtained by the proposed approach do not change with misregistration as the algorithm adopted does not take into account the spatial correspondence between pixels (and thus the alignment between images). On the contrary, all the other standard techniques are strongly affected by misregistration errors. For example, Table V shows that the HPF technique applied to misregistered images significantly degraded its performance, resulting in poor fusion results. Similar results were obtained by the other standard methods.

An additional experiment was carried out in order to assess the sensitivity of the proposed approach to the use of a symmetric or an asymmetric AR model. The quantitative evaluation of the results obtained replacing the symmetric model with an asymmetric one (in which all the 24 parameters are considered for learning the spatial details from the Pan image) are reported

TABLE V
QUANTITATIVE ASSESSMENT OF THE FUSION RESULTS PROVIDED BY THE CONSIDERED TECHNIQUES APPLIED TO THE QUICKBIRD DATA SET: (a) BAND 2, (b) BAND 3, (c) BAND 4 (THE IHS AND BT TECHNIQUES HAVE BEEN APPLIED TO BANDS 2, 3, AND 4 FALSE-COLOR COMPOSITIONS, WHILE THE OTHER TECHNIQUES HAVE BEEN APPLIED TO EACH SPECTRAL CHANNEL SEPARATELY)

| Image | Std Dev | Definition | Distortion extent | Deviation index | Corr coeff | mse |
|-------------------------|---------|------------|-------------------|-----------------|------------|-------|
| MS-Band2 | 95.098 | 26.292 | - | - | - | - |
| Bicubic (coregistered) | 74.458 | 8.785 | 37.514 | 0.107 | 0.824 | 0.022 |
| Bicubic (misregistered) | 74.145 | 8,747 | 51.268 | 0.147 | 0.645 | 0.041 |
| HPF (coregistered) | 81.762 | 21.972 | 33.893 | 0.096 | 0.859 | 0.017 |
| HPF(misregistered) | 79.487 | 20.786 | 47.83 | 0.137 | 0.705 | 0.035 |
| BT (coregistered) | 96.173 | 31.066 | 33.321 | 0.098 | 0.876 | 0.017 |
| BT (misregistered) | 97.022 | 31.001 | 36.352 | 0.108 | 0.852 | 0.021 |
| IHS (coregistered) | 90.367 | 27.304 | 30.857 | 0.090 | 0.8917 | 0.014 |
| IHS (misregistered) | 88.872 | 27.033 | 33.774 | 0.099 | 0.872 | 0.017 |
| GS(coregistered) | 78.767 | 23.327 | 31.558 | 0.091 | 0.886 | 0.015 |
| GS(misregistered) | 83.079 | 21.022 | 33.638 | 0.097 | 0.870 | 0.016 |
| Proposed (symmetric) | 88.910 | 24.310 | 27.256 | 0.078 | 0.917 | 0.011 |
| Proposed(asymmetric) | 88.765 | 22.960 | 26.580 | 0.075 | 0.920 | 0.010 |

(a)

| Image | Std Dev | Definition | Distortion extent | Deviation index | Corr coeff | mse |
|-------------------------|---------|------------|-------------------|-----------------|------------|--------|
| MS-Band3 | 82.169 | 22.579 | - | - | - | - |
| Bicubic (coregistered) | 64.933 | 7.627 | 32.222 | 0.138 | 0.834 | 0.030 |
| Bicubic (misregistered) | 64.688 | 7.639 | 44.242 | 0.191 | 0.659 | 0.057 |
| HPF (coregistered) | 70.182 | 17.577 | 28.934 | 0.123 | 0.867 | 0.024 |
| HPF(misregistered) | 68.174 | 16.186 | 41.27 | 0.178 | 0.715 | 0.0497 |
| BT (coregistered) | 78.488 | 22.509 | 25.077 | 0.108 | 0.900 | 0.019 |
| BT (misregistered) | 76.507 | 22.059 | 27.674 | 0.121 | 0.881 | 0.022 |
| IHS (coregistered) | 81.160 | 26.751 | 35.608 | 0.164 | 0.859 | 0.032 |
| IHS (misregistered) | 81.592 | 26.746 | 36.432 | 0.170 | 0.850 | 0.034 |
| GS(coregistered) | 69.243 | 21.007 | 26.994 | 0.117 | 0.890 | 0.020 |
| GS(misregistered) | 72.993 | 19.023 | 29.033 | 0.126 | 0.873 | 0.023 |
| Proposed (symmetric) | 81.848 | 23.543 | 25.018 | 0.113 | 0.916 | 0.017 |
| Proposed(asymmetric) | 80.05 | 22.06 | 22.64 | 0.098 | 0.921 | 0.015 |

(b)

| Image | Std Dev | Definition | Distortion extent | Deviation index | Corr coeff | mse |
|-------------------------|---------|------------|-------------------|-----------------|------------|--------|
| MS-Band4 | 91.441 | 25.718 | - | - | - | - |
| Bicubic (coregistered) | 72.218 | 7.534 | 35.920 | 0.122 | 0.833 | 0.020 |
| Bicubic (misregistered) | 73.490 | 7.839 | 45.490 | 0.157 | 0.705 | 0.033 |
| HPF (coregistered) | 77.209 | 18.137 | 32.700 | 0.109 | 0.861 | 0.017 |
| HPF(misregistered) | 68.174 | 16.186 | 41.27 | 0.178 | 0.715 | 0.0497 |
| BT (coregistered) | 95.814 | 28.308 | 32.951 | 0.103 | 0.880 | 0.016 |
| BT (misregisterd) | 98.443 | 29.040 | 40.262 | 0.124 | 0.830 | 0.025 |
| IHS (coregistered) | 94.504 | 27.461 | 32.275 | 0.102 | 0.878 | 0.016 |
| IHS (misregistered) | 95.785 | 27.965 | 38.742 | 0.122 | 0.834 | 0.023 |
| GS(coregistered) | 76.883 | 20.641 | 32.236 | 0.107 | 0.873 | 0.015 |
| GS(misregistered) | 80.422 | 18.561 | 32.879 | 0.110 | 0.871 | 0.016 |
| Proposed (symmetric) | 86.320 | 24.084 | 30.323 | 0.096 | 0.891 | 0.014 |
| Proposed (asymmetric) | 86.27 | 22.09 | 28.81 | 0.093 | 0.898 | 0.013 |

(c)

TABLE VI
GLOBAL QUANTITATIVE ASSESSMENT OF THE FUSION RESULTS PROVIDED BY THE CONSIDERED TECHNIQUES IN TERMS OF RASE AND ERGAS MEASURES FOR BANDS 4, 3, AND 2 OF THE QUICKBIRD DATA SET

| Image | RASE | ERGAS |
|-------------------------|-------|-------|
| Bicubic (coregistered) | 0.974 | 4.156 |
| Bicubic (misregistered) | 1.190 | 5.118 |
| HPF (coregistered) | 1.038 | 4.567 |
| HPF (misregistered) | 1.176 | 5.083 |
| BT (coregistered) | 0.682 | 2.883 |
| BT (misregistered) | 1.705 | 7.329 |
| IHS (coregistered) | 0.699 | 2.979 |
| IHS (misregistered) | 1.726 | 7.426 |
| GS (coregistered) | 0.571 | 2.385 |
| GS (misregistered) | 0.804 | 3.528 |
| Proposed (Symmetric) | 0.372 | 1.580 |
| Proposed (Asymmetric) | 0.452 | 1.896 |

in Table V [see “Proposed (asymmetric)”. The corresponding fusion result is shown in Fig. 9(b). As one can see, the differences between the symmetric and the asymmetric models on this data set are not particularly significant. For this reason, it does not seem reasonable to increase the computational load by replacing the symmetrical AR model with an asymmetric one.

Finally, we show the results of the experiments carried out using the spectral bands 3, 2, and 1 for the Quickbird data set. The obtained fused images and the related quantitative assessment are reported in Figs. 10(a)–(f) and 11, and Table VII (for simplicity, here, we only present the quantitative assessment for Pan sharpening of MS band 1). By analyzing Table VII, we can observe that, unlike the case of the ETM+ images, the quantitative results point out that both the spectral and spatial distortions obtained with the proposed technique on the channel outside the range of the Pan data slightly increases with respect to the use of spectral channels inside the range of the Pan image. However, looking at the measures given for other approaches, and taking into account the properties of the proposed approach, it may be concluded that the proposed method yielded satisfactory results.

VI. DISCUSSION AND CONCLUSION

In this paper, a novel model-based approach to multiresolution fusion of Pan and MS images has been presented. The spatial dependence in the MS images is learnt from the Pan observations by using an AR model, and a suitable regularization technique is defined to enhance the resolution of MS data.

Experimental results obtained on coregistered and misregistered MS and Pan images acquired by Landsat-7 ETM+ (1 : 2 fusion ratio) and Quickbird satellite (1 : 4 fusion ratio) sensors confirm that, in general, the proposed approach significantly increased the spectral fidelity of the fused images (with respect to other reference techniques considered), by maintaining good properties in terms of spatial enhancement. In particular, it is very effective in modeling the texture of the analyzed scene.

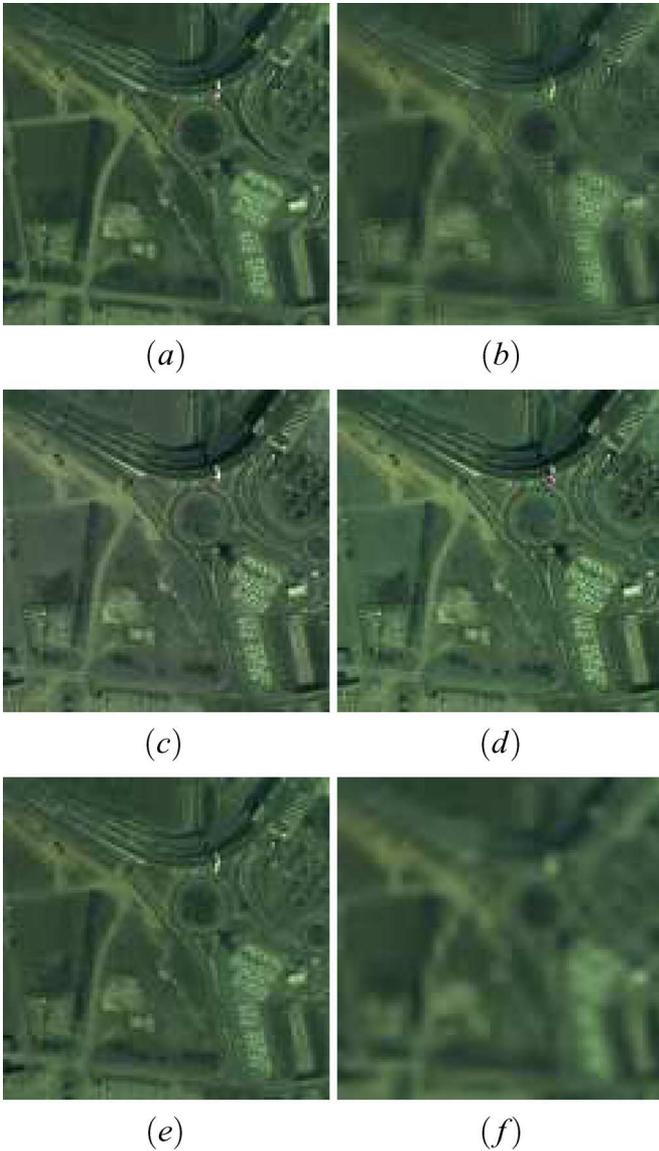


Fig. 10. True-color composite for fusion results at $2.4\text{ m} \times 2.4\text{ m}$ obtained from the degraded bands 3, 2, and 1 (Quickbird data set). (a) True bands 3, 2, and 1 image. (b) HPF fusion. (c) IHS fusion. (d) BT fusion. (e) Results of the GS method. (f) Bicubically interpolated bands 3, 2, and 1 image.

However, the main theoretical and practical advantage of the proposed method over the techniques reported in the literature is that it does not require any pixel-to-pixel correspondence (registration) between the MS and the Pan data. This avoids the introduction of artifacts due to residual misalignment between images in the fusion process. It is worth noting that this is a very important property, since it is often not possible to obtain a perfect registration between Pan and MS images (especially when Pan and MS data are acquired by different sensors). Another important intrinsic property of the proposed approach is that it can be applied to Pan and MS images acquired (also by different sensors) in slightly different geographical areas.

Although we obtained very good results on both the considered data sets, by analyzing the method from a theoretical viewpoint, we expect that, in general, it is more effective on moderately high resolution images rather than on very high resolution images, since the finer details present in high res-



Fig. 11. Fusion results at $2.4\text{ m} \times 2.4\text{ m}$ obtained from the degraded MS bands 3, 2, and 1 by using the proposed model-based approach (Quickbird data set).

TABLE VII
QUANTITATIVE ASSESSMENT OF THE FUSION RESULTS OBTAINED WITH THE CONSIDERED TECHNIQUES APPLIED TO BAND 1 OF THE QUICKBIRD DATA SET (THE IHS AND BT TECHNIQUES WERE APPLIED TO BANDS 3, 2, AND 1 FALSE-COLOR COMPOSITIONS, WHILE THE OTHER TECHNIQUES WERE ONLY APPLIED TO BAND 1)

| Image | Std Dev | Definition | Distortion extent | Deviation index | Corr coeff | mse |
|----------|---------|------------|-------------------|-----------------|------------|-------|
| MS-Band1 | 53.166 | 14.461 | - | - | - | - |
| Bicubic | 41.771 | 4.927 | 20.554 | 0.080 | 0.825 | 0.015 |
| HPF | 46.656 | 13.427 | 19.385 | 0.076 | 0.851 | 0.012 |
| BT | 61.881 | 23.036 | 46.453 | 0.197 | 0.696 | 0.047 |
| IHS | 68.230 | 26.200 | 51.866 | 0.222 | 0.669 | 0.061 |
| GS | 43.530 | 12.183 | 18.361 | 0.074 | 0.875 | 0.011 |
| Proposed | 64.152 | 22.830 | 23.919 | 0.098 | 0.850 | 0.018 |

olution data are difficult to model using AR parameterization. Nonetheless, on the considered Quickbird data set, the proposed approach resulted in very good spectral and spatial fidelity with respect to standard techniques. These performances (which are due to the stationarity of AR parameters in the considered scene) are particularly relevant and should be emphasized in the light of the aforementioned attractive properties of the technique presented.

Concerning the effectiveness of the presented algorithm on the fusion of spectral channels outside the range of the Pan image, it depends on the level of correlation of the texture or of the spatial features between the Pan and the considered MS channel.

As a final remark, it is worth noting that the quality of the fused images provided by the proposed method also depends on the stationarity of the values of AR parameters along the spatial domain of the images. For this reason, in order to increase the effectiveness of the proposed algorithm, in future developments of this paper, we plan to study the definition of a measure of stationarity for the AR parameters capable of driving a split of the considered scene into different subblocks, for which different values of the AR parameters can be computed from the Pan image. This seems particularly promising to further increase the spatial fidelity of the fused images, especially when very high resolution images are considered.

Others interesting future developments of this paper relate to: 1) the assessment of the performance of the model-based

approach in the fusion of Pan and MS images acquired by different sensors at different times (e.g., to fuse the SPOT Pan and the Landsat ETM+ MS images) and 2) the assessment of the effectiveness of the proposed approach in very critical cases in which the Pan image for the investigated region is not available, but a Pan image of similar textured (vegetation, urban structures) regions is given.

ACKNOWLEDGMENT

The Landsat-7 ETM+ image of the city of Trento was acquired in the framework of the project ASI 175/02 (funded by the Italian Space Agency) and kindly provided by R. Rigon (University of Trento, Italy). The authors are thankful to L. Alparone (University of Florence, Italy) for his useful suggestions, to L. Carlin (University of Trento, Italy) for his help and comments, and to the five reviewers for their critical yet helpful comments.

REFERENCES

- [1] L. Wald, "Some terms of references in data fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1190–1193, May 1999.
- [2] A. Farina, M. Costantini, and F. Zirilli, "Fusion of radar images: Techniques and applications," in *Invited Lecture, Colloq. Entretiens Science et Defence, topic on Le futur du radar: Une synthesis de techniques*, Jan. 24–25, 1996, pp. 285–296.
- [3] J. Hill, C. Diemer, O. Stover, and T. Udelhoven, "A local correlation approach for the fusion of remote sensing data with different spatial resolutions in forestry applications," presented at the Int. Archives Photogrammetry and Remote Sensing (ISPRS), vol. 32, Valladolid, Spain, 1999, Paper W6. Part 7.
- [4] V. Clement, G. Giraudon, S. Houzella, and F. Sandakly, "Interpretation of remotely sensed images in a context of multisensor fusion using a multispecialist architecture," *IEEE Trans. Geosci. Remote Sens.*, vol. 31, no. 4, pp. 779–791, Jul. 1993.
- [5] C. Pohl and J. L. van Genderen, "Multisensor image fusion in remote sensing: Concepts, methods and applications," *Int. J. Remote Sens.*, vol. 19, no. 5, pp. 823–854, 1998.
- [6] G. Pajares and J. M. de la Cruz, "A wavelet-based image fusion tutorial," *Pattern Recognit.*, vol. 37, no. 9, pp. 1855–1872, 2004.
- [7] A. H. J. M. Pellemans, R. W. L. Jordans, and R. Allewijn, "Merging multispectral and panchromatic spot images with respect to the radiometric properties of the sensor," *Photogramm. Eng. Remote Sens.*, vol. 59, no. 1, pp. 81–87, 1993.
- [8] R. Hayden, G. W. J. Henkel, and J. E. Bare, "Application of the IHS color transform to the processing of multisensor data and image enhancement," in *Proc. Int. Symp. Remote Sens. Arid and Semi-Arid Lands*, Cairo, Egypt, 1982, pp. 599–616.
- [9] J. Vrabel, "Multispectral imagery band sharpening study," *Photogramm. Eng. Remote Sens.*, vol. 62, no. 9, pp. 1075–1083, 1996.
- [10] H. Jing and L. Liu, "Color space conversion methods' applications to the image fusion," *Opt. Technol.*, no. 4, pp. 44–48, 1997.
- [11] 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," *Photogramm. Eng. Remote Sens.*, vol. 57, no. 3, pp. 295–303, 1991.
- [12] 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.
- [13] M. Ehlers, "Multisensor image fusion techniques in remote sensing," *ISPRS J. Photogramm. Remote Sens.*, vol. 46, no. 1, pp. 19–30, Feb. 1991.
- [14] H. Yesou, Y. Besnus, and Y. Rolet, "Extraction of spectral information from Landsat TM data and merger with SPOT panchromatic imagery—A contribution to the study of geological structures," *ISPRS J. Photogramm. Remote Sens.*, vol. 48, no. 5, pp. 23–36, Oct. 1993.
- [15] H. Li, B. S. Manjunath, and S. K. Mitra, "Multisensor image fusion using the wavelet transform," *Graph. Models Image Process.*, vol. 27, no. 3, pp. 235–244, 1995.
- [16] D. A. Yocky, "Image merging and data fusion by means of the discrete two dimensional wavelet transform," *J. Opt. Soc. Amer.*, vol. 12, no. 8, pp. 1834–1841, 1995.
- [17] E. Yu and R. Wang, "Fusion and enhancement of the multispectral image with wavelet transform," *Comput. Eng. Sci.*, vol. 23, no. 1, pp. 47–50, 2001.
- [18] J. Numez, "Multiresolution-based image fusion with additive wavelet decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1204–1211, May 1999.
- [19] Y. Chibani and A. Houacine, "Model for multispectral and panchromatic image fusion," in *Proc. SPIE Image Signal Process. Remote Sens.*, 2000, vol. 4170, pp. 238–244.
- [20] A. Garzelli, G. Bennelli, M. Barni, and C. Magini, "Improving wavelet based merging of panchromatic and multispectral images by contextual information," in *Proc. SPIE Image Signal Process. Remote Sens.*, 2000, pp. 82–91.
- [21] B. Aiazzi, L. Alparone, S. Baronti, and A. Garzelli, "Context driven fusion of high spatial and spectral resolution images based on oversampled multiresolution analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 10, pp. 2300–2312, Oct. 2002.
- [22] A. Garzelli and F. Nencini, "Interband structure modeling for panchromatic sharpening of very high-resolution multispectral images," *Inf. Fusion*, vol. 6, no. 3, pp. 213–224, 2005.
- [23] M. V. Joshi, S. Chaudhuri, and P. Rajkiran, "A learning based method for image superresolution from zoomed observations," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 35, no. 3, pp. 527–537, Jun. 2005.
- [24] —, "Super-resolution imaging: Use of zoom as a cue," *Image Vis. Comput.*, vol. 22, no. 14, pp. 1185–1196, 2004.
- [25] W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-based super-resolution," *IEEE Comput. Graph. Appl.*, vol. 22, no. 2, pp. 56–65, Mar./Apr. 2002.
- [26] D. Capel and A. Zisserman, "Super-resolution from multiple views using learnt image models," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recog.*, 2001, pp. II:627–II:634.
- [27] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 9, pp. 1167–1183, Sep. 2002.
- [28] J. Mao and A. K. Jain, "Texture classification and segmentation using multiresolution simultaneous autoregressive models," *Pattern Recog.*, vol. 25, no. 2, pp. 173–188, 1992.
- [29] R. Kashyap and R. Chellappa, "Estimation and choice of neighbors in spatial-interaction models of images," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 1, pp. 60–72, Jan. 1983.
- [30] 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, no. 1/2, pp. 4–18, Jun. 2003.
- [31] 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.
- [32] R. R. Schultz and R. L. Stevenson, "A Bayesian approach to image expansion for improved definition," *IEEE Trans. Image Process.*, vol. 3, no. 3, pp. 233–242, May 1994.
- [33] L. Bruzzone and D. Fernandez Prieto, "An adaptive semi-parametric and context-based approach to unsupervised change detection in multitemporal remote sensing images," *IEEE Trans. Image Process.*, vol. 11, no. 4, pp. 452–466, Apr. 2002.
- [34] —, "Automatic analysis of the difference image for unsupervised change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 3, pp. 1171–1182, May 2000.
- [35] Y. Yu and Q. Cheng, "MRF parameter estimation by an accelerated method," *Pattern Recognit. Lett.*, vol. 24, no. 9/10, pp. 1251–1259, Jun. 2003.
- [36] S. Z. Li, *Markov Random Field Modeling in Computer Vision*. Berlin, Germany: Springer-Verlag, 1995.
- [37] *eCognition User Guide*, Definiens Imaging GmbH, Munchen, Germany, 2005.
- [38] [Online]. Available: <http://www.gsfc.nasa.gov/IAS/handbook/handbook.htmls/chapter8/chapter8.html>
- [39] W. Cao, B. Li, and Y. Zhang, "A remote sensing image fusion method based on PCA transform and wavelet packet transform," in *Proc. IEEE Int. Conf. Neural Netw. Signal Process.*, Nanjing, China, Dec. 2003, pp. 976–981.
- [40] M. G. Audicana, J. L. Saleta, R. G. Catalán, and R. García, "Fusion of multispectral and panchromatic images using improved IHS and PCA mergers based on wavelet decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 6, pp. 1291–1299, Jun. 2004.
- [41] [Online]. Available: www.eurimage.com/products/quickbird.html



Manjunath V. Joshi (M'04) received the B.E. degree in electronics and communication engineering from Mysore University, Mysore, India, and the M.Tech. and Ph.D. degrees, from the Indian Institute of Technology, Bombay, India.

He has worked in various positions at the Gogte Institute of Technology, Belgaum, India, prior to joining the Dhirubhai Ambani Institute of Information and Communication Technology, Gujarat, India in July 2005. His research interests include image processing, signal processing, and computer vision.

He has coauthored a book titled *Motion-Free Super-Resolution* (Springer, 2005).

Dr. Joshi is a member of the Institution of Electronics and Telecommunication Engineers and the Indian Society for Technical Education.



Lorenzo Bruzzone (S'95–M'98–SM'03) received the laurea (M.S.) degree in electronic engineering (summa cum laude) and the Ph.D. degree in telecommunications from the University of Genoa, Genoa, Italy, in 1993 and 1998, respectively.

From 1998 to 2000, he was a Postdoctoral Researcher at the University of Genoa. From 2000 to 2001, he was an Assistant Professor at the University of Trento, Trento, Italy, and from 2001 to 2005, he was an Associate Professor at the same university.

Since March 2005, he has been a Full Professor of telecommunications at the University of Trento, where he currently teaches remote sensing, pattern recognition, and electrical communications. He is currently the Head of the Remote Sensing Laboratory in the Department of Information and Communication Technology, University of Trento. His current research interests are in the area of remote-sensing image processing and recognition (analysis of multitemporal data, feature selection, classification, regression, data fusion, and machine learning). He conducts and supervises research on these topics within the frameworks of several national and international projects. Since 1999, he has been appointed Evaluator of project proposals for the European Commission. He is the author (or coauthor) of more than 150 scientific publications, including journals, book chapters, and conference proceedings. He is a Referee for many international journals and has served on the Scientific Committees of several international conferences.

Dr. Bruzzone ranked first place in the Student Prize Paper Competition of the 1998 IEEE International Geoscience and Remote Sensing Symposium (Seattle, July 1998). He was a recipient of the Recognition of IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING Best Reviewers in 1999 and was a Guest Editor of a Special Issue of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING on the subject of the analysis of multitemporal remote-sensing images (November 2003). He was the General Chair and Cochair of the First and Second IEEE International Workshop on the Analysis of Multi-temporal Remote-Sensing Images. Since 2003, he has been the Chair of the SPIE Conference on Image and Signal Processing for Remote Sensing. He is an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. He is a member of the Scientific Committee of the India–Italy Center for Advanced Research. He is also a member of the International Association for Pattern Recognition and of the Italian Association for Remote Sensing (AIT).



Subhasis Chaudhuri (S'86–M'87–SM'02) was born in Bahutali, India. He received the B.Tech. degree in electronics and electrical communication engineering from the Indian Institute of Technology (IIT), Kharagpur, India, in 1985 and the M.S. degree in electrical engineering from the University of Calgary, Calgary, AB, Canada, and the Ph.D. degree in electrical engineering from the University of California, San Diego.

He joined IIT, Bombay, in 1990, as an Assistant Professor and is currently serving as Professor and Head of the Department. He has also served as a Visiting Professor at the University of Erlangen, Nuremberg, Germany, and the University of Paris XI. He is the coauthor of the books *Depth from Defocus: A Real Aperture Imaging Approach* (Springer, 1999) and *Motion-Free Super-Resolution* (Springer, 2005). He has also edited a book on *Super-Resolution Imaging* (Kluwer, 2001). His research interests include image processing, computer vision, and multimedia.

Dr. Chaudhuri is a Fellow of the Alexander von Humboldt Foundation, Germany, Indian National Academy of Engineering, National Academy of Sciences, and the IETE, India. He is the recipient of Dr. Vikram Sarabhai Research Award for the year 2001, the Swarnajayanti Fellowship in 2003, and the S.S. Bhatnagar Award in engineering sciences in 2004.