



HAL
open science

A method to better account for modulation transfer functions in ARSIS-based pansharpening methods

Pierre Massip, Philippe Blanc, Lucien Wald

► **To cite this version:**

Pierre Massip, Philippe Blanc, Lucien Wald. A method to better account for modulation transfer functions in ARSIS-based pansharpening methods. *IEEE Transactions on Geoscience and Remote Sensing*, 2011, 99, pp.800-808. 10.1109/TGRS.2011.2162244 . hal-00644244

HAL Id: hal-00644244

<https://minesparis-psl.hal.science/hal-00644244>

Submitted on 23 Nov 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A method to better account for modulation transfer functions in ARSIS-based pan-sharpening methods

Pierre Massip, Philippe Blanc and Lucien Wald

Abstract— Multispectral (MS) images provided by Earth Observation satellites have generally a poor spatial resolution while panchromatic images (PAN) exhibit a spatial resolution two or four times better. Data fusion is a mean to synthesize MS images at higher spatial resolution than original by exploiting the high spatial resolution of the PAN. This process is often called pan-sharpening. The synthesis property states that the synthesized multispectral images should be as close as possible to those that would have been acquired by the corresponding sensors if they had this high resolution. The methods based on the concept ARSIS are able to deliver synthesized images with good spectral quality but whose geometrical quality can still be improved. We propose a more precise definition of the synthesis property in terms of geometry. Then, we present a method that takes explicitly into account the difference in modulation transfer function (MTF) between PAN and MS in the fusion process. This method is applied to an existing ARSIS-based fusion method: ATWT-M3. Simulated images of the sensors Pleiades and SPOT-5 are used to illustrate the performances of the approach. Though this study is limited in methods and data, we observe a better restitution of the geometry and an improvement in all indices classically used in quality budget in pan-sharpening. We present also a means to assess the respect of the synthesis property from a MTF point of view.

Index Terms— image enhancement, modulation transfer function (MTF), multiscale techniques, remote sensing, image fusion, pan-sharpening.

I. INTRODUCTION

IMAGE fusion is currently used in various domains like medical imagery, computer vision, or remote sensing. The fusion process was defined by Wald [1] as “*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 a greater quality, although the exact definition of ‘greater quality’ will depend on the application*”. Fusion is very helpful in the analysis of the many data coming as images or series of numbers from several sensors with different characteristics. This paper focuses on the domain of

remote sensing and specifically on the synthesis of MS images at a spatial resolution higher than original by exploiting the high spatial resolution of another image.

In remote sensing several satellites provide images of two kinds altogether but with opposite characteristics in terms of spatial and spectral resolution. Multispectral (MS) images have a good spectral quality but a poor spatial resolution whereas panchromatic (PAN) images have on the contrary a high spatial resolution but with a coarser/poorer spectral quality. These two types of images allow identifying observed structures through different information by the spectral signature on the one hand and by geometrical information on the other hand. The spatial and spectral resolutions of a sensor are technically linked; a high resolution for both at the same time is hardly feasible [2]. The coarse spatial resolution of MS images is the result of a trade-off due to physical and technical reasons and constraints. The quantity of energy which arrives onto the detector is proportional to the width of its spectral range and is smaller in the MS case than in the PAN one. It is therefore necessary to increase the energy that impinges on the MS detector to obtain satisfactory signal-to-noise ratio and dynamics. Different solutions are possible. One often used is to increase the size of the MS “pixel”; it induces a degradation of the spatial resolution compared to the PAN image. In addition, if MS images were having the high resolution, the amount of data to transmit would be larger, e.g., four times more. On-board storage and data transmission to the ground are other constraints which lead to the trade-off resulting in a decrease of the spatial resolution of MS images compared to PAN images.

Data fusion is a mean to synthesize MS images at higher spatial resolution than original by exploiting the PAN high spatial resolution. This process is often called pan-sharpening. The interest of fusing both types of images has been demonstrated for long [3]-[10].

Here, the aim of fusion is to perform a high-quality transformation of the MS content when increasing the spatial resolution from the original MS images to that of the PAN image. The problem may be seen as the inference of the information that is missing to the original MS images for the construction of the MS images synthesized at a better spatial resolution [11]-[13].

The fused images should obey several properties [5] [11]-

Manuscript received April 6, 2010; revised October 25, 2011 and Mai 8, 2011.

P. Massip, Ph. Blanc and L. Wald are with the Ecole des Mines de Paris/Armines, Center for Energy and Processes, 06904 Sophia Antipolis, France (e-mail: lucien.wald@mines-paristech.fr).

[14]. The EARSeL Working Group synthesized the previous publications and proposes the consistency and synthesis properties [15]-[16]. The consistency property expresses that the fused images should offer a strong consistency with the original data set: if a fused product is down-sampled to its original low resolution, the original MS image should be retrieved. The synthesis property deals with the quality of the high-quality transformation: any synthesized image should be as identical as possible to the image that the corresponding sensor would observe with the highest spatial resolution, if existent. These properties hold for each individual modality – or spectral band– MS as well as for the multi-modality set.

Fusion methods can be classified in three families: projection-substitution, relative spectral contribution and methods belonging to the ARSIS concept [13] [16]. ARSIS is the French acronym for: “*Amélioration de la Résolution Spatiale par Injection de Structures*” (Improving spatial resolution by structure injection). Several recent methods merge principles of several families into a single method; they are called hybrid in [16].

The methods based on the ARSIS concept are inherently built to respect the consistency property [5] [11]-[13] [16]-[19]. This concept is detailed in the next section. Recently published works tend to demonstrate that, among the three families, the methods based on the ARSIS concept better synthesize images with respect to the expected spectral properties [5] [11] [13] [16]-[19]. For this reason, the present work takes place in the ARSIS concept.

However, the need to refine the spatial/geometrical quality of ARSIS-synthesized images is often underlined [5] [10] [16]. This article aims at such a refinement. Following [20], the geometry is defined as the spatial organization of the radiometry in the image. The changes in this organization can be described by a hierarchy of changes with scale (or frequencies), *i.e.*, a multiscale representation. The ARSIS concept comprises such a multiscale model (MSM) [5]; this facilitates the representation of the geometry.

Characteristics of the imaging instrument are important with respect to the restitution of the geometry of the landscape. Besides their spectral and spatial resolutions, an important characteristic of optical instruments is the impulse response which is named point spread function (PSF). The PSF has a dual function: the modulation transfer function (MTF). The MTF is the modulus of the Fourier transform of the PSF. The MTF determines the upper limits in image quality, *i.e.*, the image resolution or sharpness, and describes the image quality in terms of contrast as a function of spatial frequency [21].

The difference in sampling rates, pixel size, integration time, and spectral sensitivity, between MS and PAN sensors yields a difference in MTF. This difference must be taken into account when synthesizing MS images at the PAN high spatial resolution. Few fusion methods already include this difference in MTF through the multiscale model (MSM) and only when the value of the MTF at the Nyquist frequency is known, such as the case of Quickbird [22]-[24]. In these methods, the low pass filter of the MSM is a Gaussian. It is designed separately for each spectral band in such a way that at the Nyquist frequency its gain is equal to the value of the MTF of the corresponding band of Quickbird. These works demonstrate that taking into account the MTF improves the respect of the synthesis property and therefore the performances of the fusion methods.

We propose a method that takes the difference in MTF into account in ARSIS-based methods. It differs from previous works because it is conceived as an add-on to existing fusion methods with no need to design specific filters.

II. THE ARSIS CONCEPT

The ARSIS concept is very general and does not specify any mathematics tool. It assumes that the missing information can be synthesized thanks to the available information through the various modality and scales [5] [12].

In the following, we use the notations of [5] [16]. A defines the PAN image and B_k the k^{th} spectral band MS. N is the total number of MS bands. In the case of Quickbird images where $N = 4$, k equal to 1 designates the blue modality while k equal to 4 represents the near infrared one. Subscript ‘ $_0$ ’ is the PAN image spatial resolution res_0 , thus the original PAN image is called A_0 . Subscript ‘ $_1$ ’ is the resolution index of the original MS images res_1 , giving B_{k1} . ‘ $*$ ’ designates fused products: fused MS images at spatial resolution of PAN res_0 are called $(B_{k1})^*_0$. The issue of the synthesis of MS images to a higher spatial resolution can thus be mathematically expressed [16]:

$$(B_{k1})^*_0 = f(B_{k1}, A_0) \quad (1)$$

The aim of fusion is to perform a high-quality transformation of the MS content, when increasing the spatial resolution, from res_1 to res_0 . The problem may be seen as the inference of the information that is missing to the images B_{k1} for the construction of the synthesized images $(B_{k1})^*_0$.

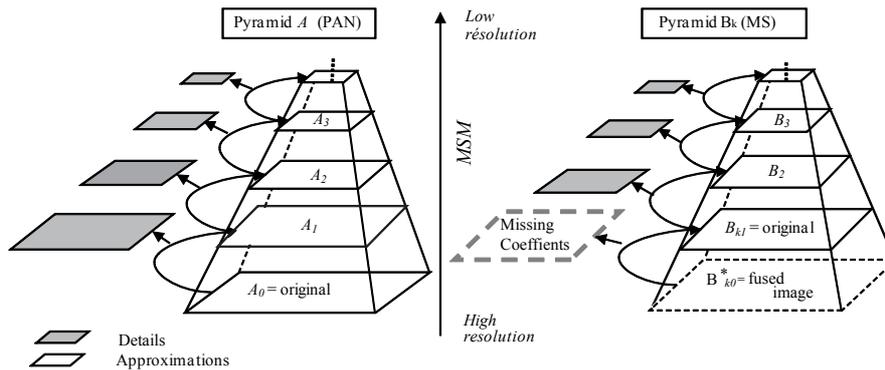


Fig. 1: Hierarchical information decomposition. MSM: multiscale model. From [16]

An important tool in the ARSIS-based methods is the MSM. It performs a hierarchical decomposition of the information of the images A and B_{kl} , represented as pyramids in Fig. 1. When climbing the pyramid (the analysis step), the successive approximations have coarser and coarser spatial resolution. The analysis of an image produces an approximation of the image at a coarser scale and details that describe the difference between the approximation and the finer image. Recent methods exploit high-pass filtering, Laplacian pyramids, decimated or undecimated wavelet transforms [5] [10]-[12] [16]-[19]. The selected MSM must be reversible.

Fig. 2 illustrates the three models that are found in ARSIS-based methods [12]:

- the MSM and its reverse MSM^{-1} ,
- the inter-modality model (IMM). It relates the MS and PAN images at a given scale, and allows to make approximation of one image from another and *vice-versa*,
- the high resolution inter-modality model (HRIMM). The HRIMM computes (infer) the missing information by exploiting the IMM and the available information from the image A_0 . Combined with MSM^{-1} , it performs the synthesis of the high spatial resolution MS images $(B_{kl})^*_{0}$.

An ARSIS-based method is composed of one model of each type.

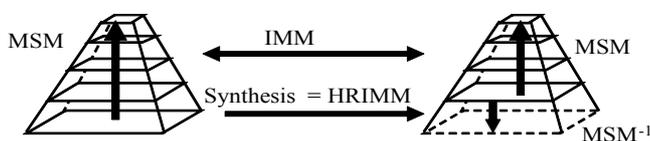


Fig. 2: Models in ARSIS. See text for more explanation. MSM: multiscale model; IMM: inter-modality model; HRIMM: high resolution inter-modality model. (C. Thomas, personal communication)

III. METHODOLOGY

As said before, the ARSIS-based methods are inherently built to respect the consistency property. Consequently, the present work focuses on the synthesis property. This property has been expressed in general terms [11] [13] [15]-[16]. We propose here a precise definition of the synthesis property in

terms of geometry: the restitution of the spatial organization of the radiometry in the observed landscape by the synthesized image should be as identical as possible to the organization that the corresponding sensor would observe with the highest spatial resolution, if existent.

The work presented here aims at including the difference in MTF between PAN and MS images into an existing ARSIS-based fusion method in order to demonstrate that taking into account this difference leads to a better respect of the synthesis property from the geometrical point of view without degrading the quality of the other aspects of the synthesized images.

The fusion method used in this demonstration is the method ATWT-M3, where ATWT denotes the MSM and means “à trous” (with holes) wavelet transform, which is an undecimated transform, and M3 is the inter-modality model described by [12]-[13] [19]. The HRIMM is identical to the IMM as in most published works. This method has been shown to provide good results in most cases. It is rather simple and allows the assessment of the impact resulting from the MTF adaptation. While the results will not be the same than those obtained here, our method for accounting for the difference in MTF may easily apply to other ARSIS-based methods.

The MTFs for the four MS channels of the Pleiades sensor, and that for the PAN, are drawn in Fig. 3 as a function of the spatial frequency normalized by the sampling frequency of the PAN image. Consequently the relative Nyquist frequency for PAN is 0.5 and 0.125 for MS. The spatial resolution of the PAN image (70 cm) is four times better than that of the MS (280 cm). The MTFs mainly account for detector performances and blurring since the MTF for each MS is almost equal to zero for relative frequencies close to 0.25. It means that one cannot distinguish details less than 280 cm in size in the MS images. On the contrary, the MTF of PAN is large –equal to 0.4– for this frequency and such details are highly visible. The difference in MTF between low and high spatial resolution images is evidenced in this figure: the MTF of MS has a sharper decrease with frequency than that of PAN.

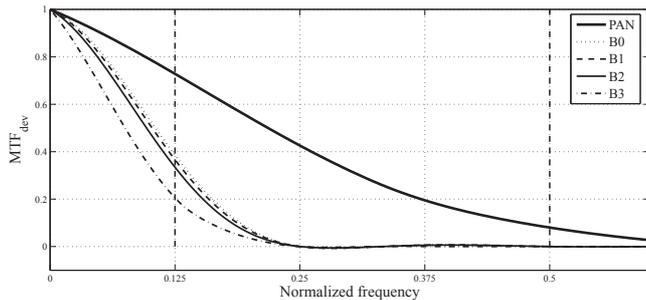


Fig. 3: Modulation transfer function for Pleiades MS and PAN images

According to the synthesis property, the synthesized images should exhibit the typical MTF of real images at this resolution. However, without modification of the MTF during the fusion process, the resulting MTF exhibits a discontinuity as shown in broad dash in Fig. 4. In this figure are drawn the schematic representation of the MTF for *a)* the low resolution, *b)* the high resolution, and *c)* the synthesized image without modification of the MTF. The curves for the MTF at low and high resolution are similar to those for MS and PAN in Fig. 3.

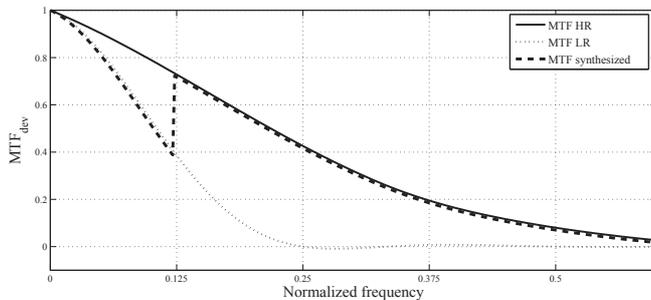


Fig. 4: Schematic representation of the MTF for synthesized MS, MS (LR) and PAN images (HR)

One can see the discontinuity in the MTF of the synthesized image for the relative frequency equal to 0.125. Below this value, the MTF is similar to that of the MS at low resolution. Above this value, the MTF is similar to that of a high resolution image. Such a discontinuity should not exist if the synthesis property were to be respected. Its existence explains partly several artifacts observed in synthesized images. It also illustrates the expected benefit of taking into account the difference in MTF during the fusion process. In an illustrative way-of-speaking we can say that the solution consists in “raising” the MTF of the MS frequencies in the range $[0, 0.125]$ so that it is close to that of a high resolution image. Doing so provides a MTF closer to a “real” MTF and similar contrast in the image for a same frequency without considering its origin: PAN or MS. Actually, it is an adaptation of the original low-resolution MTF of MS images to a high-resolution MTF for the synthesized images. This is why we will use the term “MTF adaptation”.

The method ATWT-M3 is as follows [12]. As the MSM is a non decimated wavelet transform, it implies a preliminary zoom/resampling of the original MS images B_{kl} to the resolution $res0$ by a spline technique [25]. The MSM is applied to these resampled images as well as to the A_0 image firstly at the resolution $res0$ and then at $res1$. For each

modality k and for the scale $res1$, an affine function between the PAN and MS planes of wavelet coefficients is determined by a least-square fitting. This is the IMM model. The HRIMM is identical to the IMM and the missing wavelet coefficients at high resolution for B images are computed from those of the image A . Finally, the inverse wavelet transform is applied to obtain the synthesized MS images $(B_{kl})^*_{0}$.

We derive a new method from this standard method, called ATWT-M3-MTFadapted. It is identical to ATWT-M3, except that we perform a transformation of the MS images before and after the preliminary resampling of the MS image. The MTF adaptation starts with a deconvolution of the original MS images B_{kl} . The MTF of a sensor is the sum of two MTFs: the optical one and the detector one. In a first approximation the optical MTF is the same for each spectral band in most of the current sensors. Therefore, the optical MTF is not taken into account as it does not change between MS and PAN and between MS bands. The deconvolution kernel corresponds to the MTF of the detector. [26] proposes a method for the assessment of the MTF of the detector and its modeling. After the deconvolution, the image has only the optical MTF and is resampled to the resolution of PAN. Then we convolve the resampled image with the MTF of the PAN detector. Eventually, the method ATWT-M3 is applied. Fig. 5 provides a diagram that illustrates the process applied to the MS image to obtain the MTF adapted MS image and the corresponding MTF.

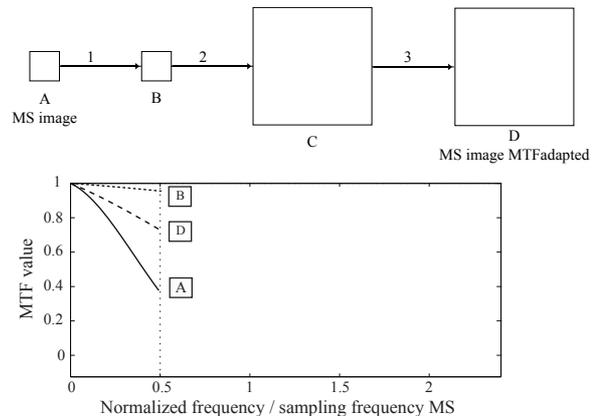


Fig. 5: Diagram representation of the process of MTF modification with the corresponding MTF curves, for each step

- 1: deconvolution of A with the low resolution detector MTF
- 2: upsampling of image B with a ratio 4 (spline function)
- 3: convolution of C, with the high resolution detector MTF

This approach to the MTF adaptation is the main innovation proposed in this article and is a means for accounting for differences in MTF, as demonstrated in the following. The approach is simple because it does not necessitate the knowledge of the global MTF for each band and the PAN. It only needs the MTF of the MS and PAN detectors.

When applied to real images, system level MTF information is required, including possible image enhancement post-processing.

This information can be given by the image provider or estimated on the images thanks to dedicated algorithms mainly based on natural or artificial patterns [27].

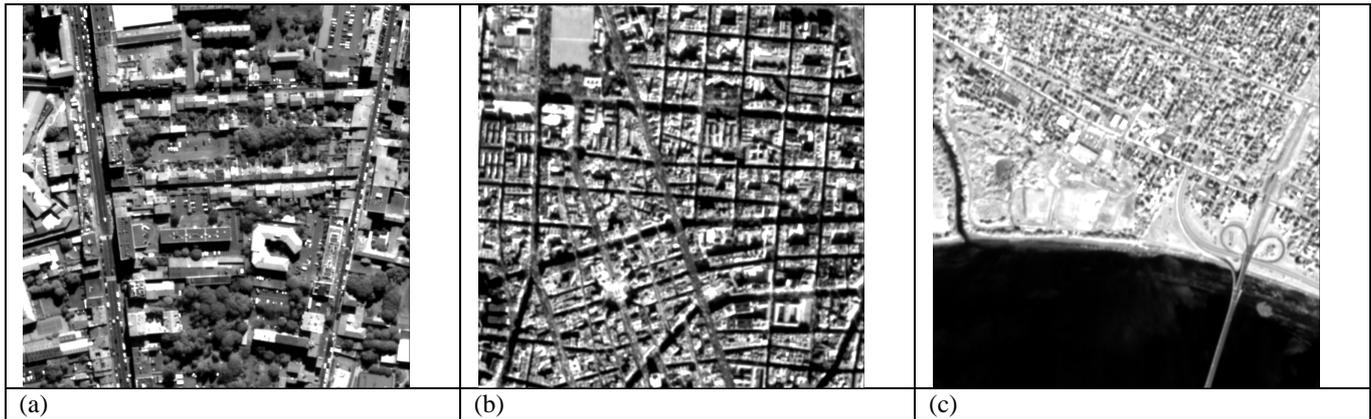


Fig. 6: Excerpt of the data sets. a) Toulouse (Pleiades simulation), b) Madrid (SPOT-5 simulation), c) Fredericton (SPOT-5 simulation)

IV. DATA SETS

Urban areas offer a large variety of spectral signature, structures and size in remote sensing imagery. Such areas evidence the performances of the fusion methods [11] [15] [22]. The data sets chosen are images of the cities of Toulouse (France), Madrid (Spain), and Fredericton (Canada).

Actually, we have also opted for simulation of images of present or to-come sensors: SPOT-5 and Pleiades. In doing so, we can control all parameters of images and we can be more conclusive in our discussion. We particularly control the properties of the MTFs and we take care that the difference in MTF between PAN and MS images is the same than in real cases. Another advantage of simulated images is that the MS images are available at both high and low spatial resolutions. This allows comparison and assessment of performances of the modified method compared to the original one. As the aim of our paper is to demonstrate the interest of taking the difference in MTF into account, our conclusions will still be general and applicable to real cases, including other sensors. The original data sets used to simulate Pleiades and SPOT-5 images are from respectively the airborne sensor Pelican and the spaceborne sensor Quickbird (Table 1).

	PAN	Blue	Green	Red/ yellow	NIR
Pelican	570-720	430-550	490-610	600-720	750-950
Pleiades	500-850	430-550	490-610	600-720	750-950
Quickbird	450-900	450-520	520-600	630-690	760-900
SPOT-5	480-710	--	500-590	610-680	780-890

Table 1: Spectral bands (in nm) for Pelican, Pleiades, Quickbird and SPOT-5

The Pelican images have been acquired with a resolution of 35 cm in all bands by the French space agency CNES. The bandwidth of the PAN modality of Pelican is [510-720] nm and does not exactly match the bandwidth of the actual Pleiades instrument which is [500-800] nm (Table 1). Pleiades images of the city of Toulouse (France) have been simulated at 70 cm and 280 cm for all bands by CNES from the Pelican images [2] –no information is available on the date of acquisition, it ranges between 1999 and 2002.– The filters

used to downsample images from 35 cm to 70 cm or 280 cm are the typical MTFs of the Pleiades satellite. The original image is firstly low-pass filtered by the MTF. Then the filtered image is resampled.

As a result, we have a PAN image at 70 cm and four MS images at 280 cm (Table 2). In addition, we have four MS images at 70 cm that will be used as reference images (B_{k0}) for assessing the performances of the method. Fig. 6 exhibits excerpts of these simulated images.

	Pelican	Simulated Pleiades	Quickbird	Simulated SPOT-5	Simulated SPOT-5 (supermode)
PAN	35 cm	70 cm	70 cm	5 m	2.5 m
MS	35 cm	280 cm	280 cm	10 m	5 m
MS reference	--	70 cm	--	5 m	2.5 m

Table 2: Spatial resolution of the acquired and simulated images

The signal-to-noise ratio (SNR) of the Pelican and the Pleiades images is very large; these images offer a very high radiometric quality. Nevertheless, we have simulated noisy Pleiades images to assess the improvements brought by the MTF adaptation in such conditions. Various levels of noise were made to test the robustness of the method with noisy images. The noise added to images corresponds to the nominal SNR multiplied by respectively 1, 0.75, and 0.5. The case with a SNR half the nominal SNR is quite extreme but is used to demonstrate the robustness of the MTF adaptation with respect to noise.

The Quickbird images have been acquired over Madrid on 29 January 2002 and over Fredericton on 21 August 2002. We simulated SPOT-5 images from these Quickbird images. The simulated images have the typical spatial resolution and difference in MTF of SPOT-5 images. Initially, we thought of deconvolving the Quickbird images prior to the application of the typical MTF of SPOT-5. However, an analysis of the Quickbird images indicates that they have been treated by a deblurring processing. In this paper the most important property in the simulation is the difference in MTF between PAN and MS and not the exact value of the MTF. Therefore, we have considered that the images have already been MTF-

restored and did not try "to remove" the MTF of Quickbird. The MTFs of SPOT-5 are applied to the Quickbird images and the resulting images are resampled to 5 m and 10 m, and to 2.5 m and 5 m for the SUPERMODE [28]. We took great care to preserve the difference in MTF between modalities that is typical to SPOT-5 images.

The spectral bands for Quickbird and SPOT-5 are not the same (Table 1). As the spectral specifications are not the issue of this paper, we did not try to match the spectral bands and sensitivities of Quickbird on those of SPOT-5. Consequently, images that simulate SPOT-5 images keep the spectral characteristics of Quickbird images.

There is a large difference in resolution between Pleiades and SPOT-5: 70 cm versus 5 m (2.5 m for SUPERMODE). We expect that the benefit of the MTF adaptation increases as the spatial resolution increases. Nevertheless, we made the SPOT5 simulation to show that the MTF adaptation offers benefit at high resolution and not only at very high resolution. Note that the ratio of resolution is 4 in the cases of Pleiades and SPOT-5 SUPERMODE, and 2 in the case of SPOT-5. Table 2 summarizes the spatial resolution of the various bands for each case.

V. RESULTS

The synthesized images $(B_{kl})^*_o$ obtained by the two methods: ATWT-M3 and ATWT-M3-MTFadaptd are compared to the reference images B_{ko} . As recommended by several authors [10]-[11] [13]-[15], we perform statistical comparisons and a visual analysis of the possible discrepancies and visual artifacts.

A. Statistical results

Quality budgets are obtained by the very same tool than that used for the IEEE data fusion contest 2006 [10] with the following monomodal indices: relative bias (biasRel), relative difference in variance in percent (diffVarRel), relative standard deviation (σ Rel), and correlation coefficient (cc), and the multimodal indices: ERGAS and mean SAM. We add the correlation coefficient of high frequency (ccHF) which is the correlation coefficient between the first wavelet planes of the reference and synthesized images [29]; it measures the performance in synthesizing the highest frequency, *i.e.* the smallest details. We also computed the multimodal indices QNR [30] and Q4 [31]; the values obtained for each method were close and these indices were not discriminatory in this study.

Tables 3 to 6 report the quality budget for each case for images of 512 x 512 pixels. Ideal values are presented for each index.

Modalities	Indices	Ideal value	ATWT-M3	ATWT-M3-MTFadaptd
Blue	biasRel (%)	0	0.00	0.01
	diffVarRel (%)	0	14	3
	σ Rel (%)	0	9	7
	cc	1	0.9833	0.9875
	ccHF	1	0.9203	0.9224
Green	biasRel (%)	0	0.00	0.01
	diffVarRel (%)	0	16	4
	σ Rel (%)	0	9	6
	cc	1	0.9875	0.9921
	ccHF	1	0.9605	0.9628
Red	biasRel (%)	0	0.00	0.01
	diffVarRel (%)	0	16	6
	σ Rel (%)	0	9	6
	cc	1	0.9905	0.9953
	ccHF	1	0.9690	0.9734
NIR	biasRel (%)	0	0.02	0.02
	diffVarRel (%)	0	34	27
	σ Rel (%)	0	14	11
	cc	1	0.9669	0.9793
	ccHF	1	0.8783	0.8886
Multimodal indices	ERGAS	0	2.6	2.0
	mean SAM	0	3.1	2.9

Table 3: Quality budget for fusion of simulated Pleiades images of Toulouse

Modalities	Indices	Ideal value	ATWT-M3	ATWT-M3-MTFadaptd
Green	biasRel (%)	0	0.00	0.00
	diffVarRel (%)	0	22	10
	σ Rel (%)	0	3	2
	cc	1	0.9799	0.9863
	ccHF	1	0.9078	0.9148
Red	biasRel (%)	0	0.00	0.00
	diffVarRel (%)	0	23	10
	σ Rel (%)	0	4	3
	cc	1	0.9826	0.9889
	ccHF	1	0.9339	0.9403
NIR	biasRel (%)	0	0.00	0.00
	diffVarRel (%)	0	20	9
	σ Rel (%)	0	5	4
	cc	1	0.9858	0.9919
	ccHF	1	0.9386	0.9471
Multimodal indices	ERGAS	0	1.1	0.8
	mean SAM	0	0.8	0.7

Table 4: Quality budget for fusion of simulated SPOT-5 supermode images of Madrid

Modalities	Indices	Ideal value	ATWT-M3	ATWT-M3-MTFadaptd
Green	biasRel (%)	0	-0.01	0.01
	diffVarRel (%)	0	12	7
	σ Rel (%)	0	5	5
	cc	1	0.9780	0.9828
	ccHF	1	0.7413	0.7582
Red	biasRel (%)	0	-0.01	0.00
	diffVarRel (%)	0	10	6
	σ Rel (%)	0	7	6
	cc	1	0.9830	0.9872
	ccHF	1	0.7969	0.8110
NIR	biasRel (%)	0	-0.02	0.03
	diffVarRel (%)	0	5	3
	σ Rel (%)	0	9	7
	cc	1	0.9923	0.9947
	ccHF	1	0.8929	0.9035
Multimodal indices	ERGAS	0	2.0	1.7
	mean SAM	0	1.9	1.6

Table 5: Quality budget for fusion of simulated SPOT-5 supermode images of Fredericton

The bias for both methods and all cases is very close to the ideal value; this is characteristic of the methods based on the ARSIS concept [10] [12]-[13] [16]. The bias is less than 1 digital count by far.

For all indices, all bands and all cases, the values obtained for ATWT-M3-MTFadapted are closer to the ideal values than for ATWT-M3. The improvement is the most significant for the relative difference in variance (diffVarRel). This index measures the lack or excess of injection of information at high frequencies. If it is positive, too many details are injected. Here the MTF adaptation reduces this excess.

The values of the correlation coefficient (cc) are already close to the ideal value. Thus, only little improvements are observed.

As for the correlation of high frequencies (ccHF), little improvement is also observed between the two methods. One may note that these coefficients are high in the case of Pleiades but low for the two SPOT-5 cases. Reasons are unknown; this may be partly due to a greater difficulty in synthesizing 5-m details from 10 m compared to 70-cm details from 280 cm given this type of landscape.

Improvement is also observed for multimodal indices. This was unexpected because the MTF adaptation implies especially contrast and sharpness. There is no process designed to improving spectral quality. An explication could be that the improvement of the monomodal quality of each band implies also an improvement of the spectral quality which is observed in multimodal indices. Another explication could be that the MTF adaptation enhances the contrast in high frequencies that are not necessarily present in the same proportion in each band. The MTF adaptation provides a better representation of the contrast for each band in the first stage of the fusion process: the MSM (Fig. 2). This allows the IMM and therefore the HRIMM to be more precise in the quantity of information to inject in the synthesized image.

B. The case of "noisy" images

Table 6 gives the quality budget for the fusion of noisy Pleiades images by ATWT-M3 and ATWT-M3-MTFadapted, where the original SNR was divided by two. The comparison between the two methods leads to the same conclusions than for the noiseless images: the MTF adaptation improves the performances of the ATWT-M3 method, even in the case of SNR divided by two compared to the nominal one.

The deconvolution amplifies the noise. Our MTF adaptation makes a deconvolution and a convolution with another MTF kernel. The entire process likely amplifies less the noise than a single deconvolution. Table 7 presents the level of SNR for images synthesized with ATWT-M3 and ATWT-M3-MTFadapted for various simulations of noise. The results confirmed the fact that the MTF adaptation amplifies noise: in each case the SNR is higher for ATWT-M3. Visually the improvement of sharpness obtained by ATWT-M3-MTFadapted is more important than the diminution of the SNR in comparison with ATWT-M3, in full agreement with the improvement in quality budget. But it is still a disadvantage if synthesized images are used for an application that is sensible to noise.

Modalities	Indices	Ideal value	ATWT-M3	ATWT-M3-MTFadapted
Blue	biasRel (%)	0	0.00	0.01
	diffVarRel (%)	0	14	3
	σRel (%)	0	9	8
	cc	1	0.9824	0.9863
	ccHF	1	0.8831	0.8865
Green	biasRel (%)	0	0.01	0.01
	diffVarRel (%)	0	15	4
	σRel (%)	0	9	7
	cc	1	0.9867	0.9910
	ccHF	1	0.9203	0.9242
Red	biasRel (%)	0	0.01	0.05
	diffVarRel (%)	0	16	6
	σRel (%)	0	10	7
	cc	1	0.9898	0.9944
	ccHF	1	0.9277	0.9335
NIR	biasRel (%)	0	-0.03	0.00
	diffVarRel (%)	0	34	26
	σRel (%)	0	14	11
	cc	1	0.9661	0.9779
	ccHF	1	0.8430	0.8531
Multimodal indices	ERGAS	0	2.6	2.0
	mean SAM	0	3.2	3.1

Table 6: Quality budget for fusion of simulated noisy Pleiades images of Toulouse

	nominal SNR			0.5 x nominal SNR		
	SNR images before fusion	SNR synthesized images		SNR images before fusion	SNR synthesized images	
		ATWT-M3	ATWT-M3-MTFadapted		ATWT-M3	ATWT-M3-MTFadapted
PAN	147	--	--	73.5	--	--
B0	130	120	93	65	60	46
B1		112	90		56	44
B2		102	84		51	41
B3		136	102		68	51

Table 7: Influence of MTF adaptation on the SNR of the synthesized image of Toulouse

C. Normalized deviation to the MTF of the reference image

All these indices demonstrate improvements brought by the MTF adaptation. Nevertheless, they do not inform on the aspect of the MTF of the synthesized images. Consequently, we perform a comparison of MTF between the reference and the synthesized images in order to assess if the adaptation of the MTF meets the expectations. Using the MTF in the quality assessment has been considered in [32]. A measure was proposed based on norms Chi2 and L2 of the differences between MTFs. [32] reported that one drawback was the inability to discriminate between over- and under-estimation of MTFs. Our proposed measure overcomes this drawback.

Here, the comparison is based on power and cross-power spectral densities (resp. PSD and cross-PSD) of images in the Fourier domain [33]. For each band k , we compute the cross-psd of the synthesized image $(B_{k1})^*_0$ and the corresponding reference image B_{k0} . We then divide this cross-psd by the PSD of B_{k0} :

$$(MTF_{dev})_k = cross-PSD(B_{k0}, (B_{k1})^*_0) / PSD(B_{k0}) \quad (2)$$

The quantity $(MTF_{dev})_k$ is a normalized deviation to the MTF of the reference image B_{k0} . It denotes the behavior of the

MTF of the synthesized image $(B_{kl})^*_0$ compared to that of B_{k0} , as a function of the relative frequency. The ideal result is a constant equal to 1 for all frequencies. Values smaller than 1 indicate a lack of injected information for a particular frequency and values greater than 1 an excess of information. Figures 7 and 8 display the results for the green band of respectively Pleiades and SPOT-5.

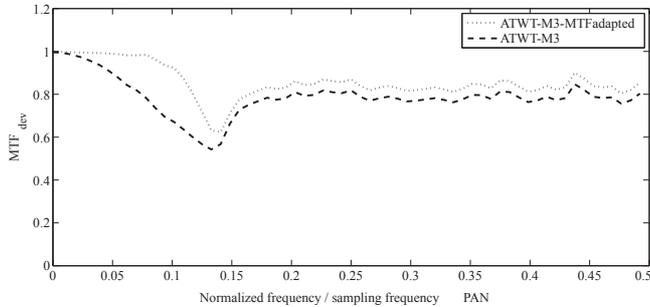


Fig. 7: Modulation transfer function of the images ATWT-M3 and ATWT-M3-MTFadapted relative to the MTF of Pleiades

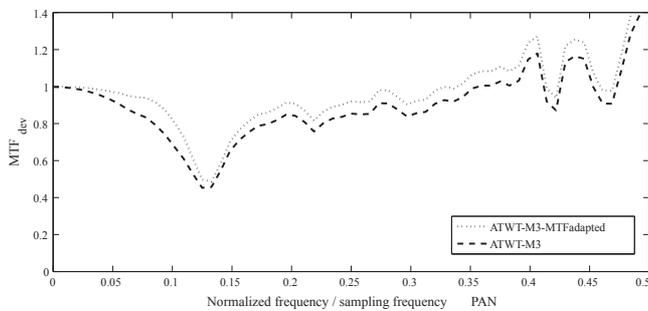


Fig. 8: Modulation transfer function of the images ATW-M3 and ATWT-M3-MTFadapted relative to the MTF of SPOT-5

In both graphs, MTF_{dev} is closer to the ideal value 1 for ATWT-M3-MTFadapted than for ATWT-M3 for all frequencies. The improvement in geometry brought by the MTF adaptation is demonstrated in this case.

We can observe according to MTF_{dev} that the improvement is less important for SPOT-5 images than for Pleiades images.

This was expected as discussed earlier.

One may note in this figure that MTF_{dev} exhibits a trough between 0.1 and 0.15. The origin of the trough is a lack of coherence between the frequencies of both images $(B_{kl})^*_0$ and B_{k0} . Tests show that it is mostly due to the combination of the preliminary resampling of the MS images because it changes the highest frequencies of the original MS images around the normalized frequency 0.125, and of the properties of the filter of the MSM. We have also observed that correcting for aliasing in MS images has a beneficial effect on this trough and partly remove it. Figures 7 and 8 clearly show that the drawback is reduced by the MTF adaptation. Nevertheless, there is room for improvements that can be brought by a smarter resampling, aliasing correction, and a better combination of the three models: MSM, IMM and HRIMM, especially taking into account the discrepancies in remote sensing physics between the PAN and MS images [16].

D. Visual analysis

Statistical results show that *i*) there is an improvement brought by the MTF adaptation, and *ii*) the improvement is more important for Pleiades than for SPOT-5. This is confirmed by visual observations.

Fig. 9 exhibits an excerpt of the red band of the simulated Pleiades for Toulouse: (a) without MTF adaptation, (b) with MTF adaptation, and (c) the reference image. The same look-up table is used for the three images. This excerpt has been selected as it fully illustrates the benefit of the MTF adaptation. Image (c) is the objective to reach. On a whole (b) is closer to (c) than (a). For example, the contour line of the white building is sharper in (b) than in (a). This is true for many other buildings, streets and cars. Several regularly-spaced windows are visible on the roof of this building in the reference image (c). Many of them can be distinguished in (b) due to the MTF adaptation whereas they are quite invisible in (a).

This holds also for color composites and here again, we find an improvement when adapting the MTF.

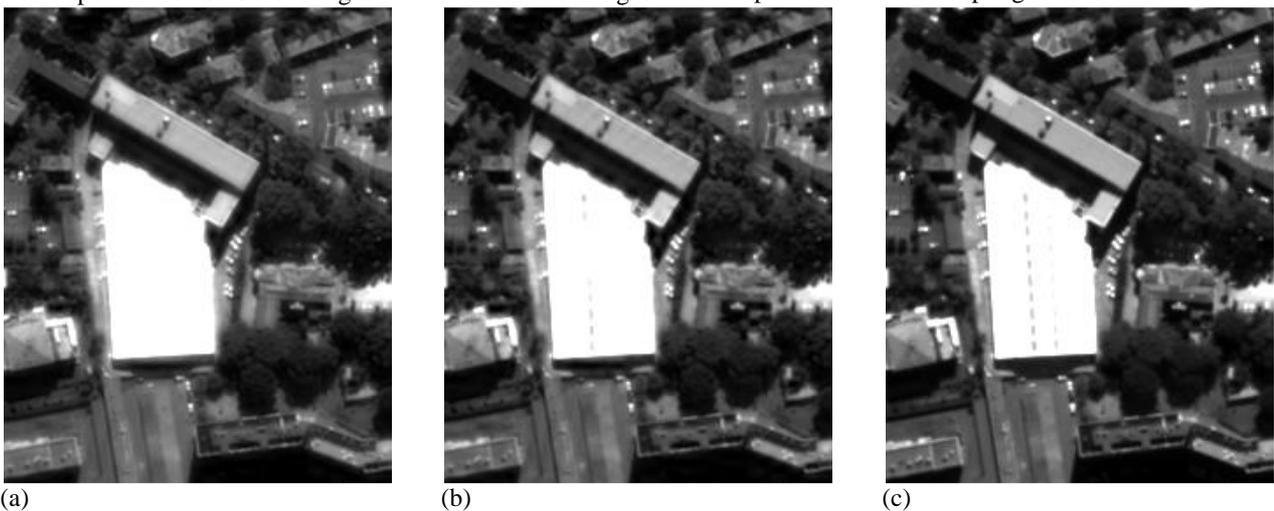


Fig. 9: Excerpt of Pleiades images over Toulouse, France. Red band synthesized with a) ATWT-M3, b) ATWT-M3-MTFadapted, c) reference image. Copyright CNES 2000.

VI. DISCUSSION

A new measure: the MTF normalized deviation MTF_{dev} , has been proposed to evaluate the performances of a fusion method with respect to the synthesis property expressed in terms of geometry. It highlights strengths and weaknesses in the MTF of the synthesized image. This measure has been used here in the case of simulation of images.

It could be used in the general case where no reference image is present at high resolution to assess the synthesis property. In that case, the protocol of [11] and [15] proposes to perform the fusion process at a scale coarser than the spatial resolution of the original MS images. Thus, the original images B_{kl} become references against which the synthesized images $(B_{kl})^*_l$ are compared. One conclusion of our work is that the degradation of resolution should take the MTF into account in an accurate way as done in the pioneering work [24].

We have demonstrated that taking into account the difference in MTF between MS and PAN images leads to synthesized images of better quality. This work can be applied to other concepts and not only to ARSIS [16]. It could be interesting to observe the difference of behavior of this technique depending on the concept it is applied on. The MTF normalized deviation is one means of observation.

We have shown how the difference in MTF between low and high resolution can be taken into account into a "standard" fusion method based on the ARSIS concept. We have observed for three cases a better restitution of the geometry and an improvement in all indices classically used in quality budget in pan-sharpening. These results demonstrate *i*) the benefit of taking into account the MTF in the fusion process and *ii*) the value of our method for MTF adaptation.

Improvement is also observed in the case of noisy images, though the SNR is decreased by the MTF adaptation.

Several methods improve contrast by injecting more details when the correlation is high between the MS and PAN images. [34] computes the correlation between the MS and PAN images at the resolution of the original MS images. The methods ATWT-RWM [5] and ATWT-M3-Sharpener [35] compute the correlation between the wavelet planes instead of the context. This increase in contrast yields an improvement of the MTF. However, the MTF is not improved in an homogeneous manner with frequency and the results depends on the correlation, on the way to compute the correlation, and on the manner to use it to increase the injection of details. Our calculations, not reported here, show that the association of the MTF adaptation with these methods is less profitable than when it is applied to a global method like ATWT-M3 presented here. By this improvement in contrast these methods improve the MTF of the fused image but the improvement is not effective if structures are not correlated. Our manner to improve the MTF is global and yields a better homogeneity of the fused image with respect to geometry.

We also evaluate the impact of the degree of the spline function used to resample images. We observed that the higher the degree, the better the MTF of the fused image. The

difference is noticeable for low degree and becomes negligible when the degree is greater than 4.

We show that pansharpening methods benefit from the use of MTF. Reference [36] also shows that quality assessment benefits from the consideration of the MTF. The pansharpening context (method and quality assessment) seems to be inseparable from the use and consideration of the MTF. Though this study is quite limited in methods and data, the present results are encouraging and may constitute a new way to improve the restitution of geometrical features by already efficient fusion methods.

VII. ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers whose comments helped in improving the clarity and quality of this paper.

REFERENCES

- [1] L. Wald, "Some terms of reference in data fusion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 3, pp. 1190-1193, 1999.
- [2] F. Laporterie-Déjean, H. De Boissezon, G. Flouzat, and M.-J. Lefèvre-Fonollosa, "Thematic and statistical evaluations of five panchromatic/multispectral fusion methods on simulated PLEIADES-HR images," *Information Fusion*, vol. 6, pp. 193-212, 2005.
- [3] A. Filippidis, L.C. Jain, and N. Martin, "Multisensor data fusion for surface land-mine detection," *IEEE Transactions on Systems, Man and Cybernetics, Part C*, vol. 30, no. 1, pp. 145-150, 2000.
- [4] P.S. Huang, and T. Te-Ming, "A target fusion-based approach for classifying high spatial resolution imagery," *In IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data*, 27-28 Oct. 2003, pp. 175-181, 2003.
- [5] T. Ranchin, B. Aiazzi, L. Alparone, S. Baronti, and L. Wald, "Image fusion-the ARSIS concept and some successful implementation schemes," *ISPRS Journal of Photogrammetry & Remote Sensing*, doi 10.1016/S0924-2716(03)00013-3, Elsevier (Eds.), vol. 58, no. 1-2, pp. 4-18, 2003.
- [6] L. Wald, and J.-M. Baleynaud, "Observing air quality over the city of Nantes by means of Landsat thermal infrared data," *International Journal of Remote Sensing*, vol. 20, no. 5, pp. 947-959, 1999.
- [7] I. Couloigner, T. Ranchin, V.P. Valtonen, and L. Wald, "Benefit of the future SPOT 5 and of data fusion to urban mapping," *International Journal of Remote Sensing*, vol. 19, no. 8, pp. 1519-1532, 1998.
- [8] P. Terretaz, "Comparison of different methods to merge SPOT P and XS data : Evaluation in an urban area," In: Gudmansen, P. (Ed.), *Proceedings of the 17th Symposium of EARSeL, Future Trends in Remote Sensing*, Lyngby, Denmark, 17-20 June. A.A. Balkema, Rotterdam: 435-445, 1997.
- [9] J. A. Malpica, "Hue adjustment to IHS Pan-sharpened Ikonos imagery for vegetation enhancement," *IEEE Geoscience and Remote Sensing Letters*, doi 10.1109/LGRS.2006.883523, vol. 4, no. 1, pp. 27-31, 2007.
- [10] L. Alparone, L. Wald, J. Chanussot, C. Thomas, P. Gamba and L.M. Bruce, "Comparison of pansharpening algorithms: outcome of the 2006 GRS-S data fusion contest," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no 10, 3012-3021, doi: 10.1109/TGRS.2007.904923, 2007.
- [11] L. Wald, T. Ranchin, and M. Mangolini, "Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images," *Photogrammetric Engineering and Remote Sensing*, vol. 63, no. 6, pp. 691-699, 1997.
- [12] T. Ranchin, and L. Wald, "Fusion of high spatial and spectral resolution images: the ARSIS concept and its implementation" *Photogrammetric Engineering and Remote Sensing*, vol. 66, no. 1, pp. 4-18, 2000.

[13] L. Wald, *Data Fusion: Definitions and Architectures. Fusion of Images of Different Spatial Resolutions*, Presses de l'Ecole (Eds.), MINES ParisTech, Paris, France, ISBN 2-911762-38-X, 200 p., 2002.

[14] J. Li, "Spatial quality evaluation of fusion of different resolution images," *In International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Amsterdam, The Netherlands, vol. 33, B2-2, pp. 339-346, 2000.

[15] C. Thomas, and Wald L., "Assessment of the quality of fused products," *In Proceedings of the 24th EARSeL Symposium "New Strategies for European Remote Sensing"*, 25-27 May 2004, Dubrovnik, Croatia, Oluc (Eds.), Millpress, Rotterdam, ISBN 90 5966 003 X, pp. 317-325, 2005.

[16] C. Thomas, T. Ranchin, Wald L., and J. Chanussot, "Synthesis of multispectral images to the high spatial resolution: a critical review of fusion methods based on remote sensing physics," *IEEE Trans. Geoscience and Remote Sensing*, vol. 46, no. 5, pp. 1301-1312, doi 10.1109/TGRS.2007.912448, 2008.

[17] D. A. Yocky, "Multiresolution wavelet decomposition image merger of Landsat Thematic Mapper and SPOT panchromatic data," *Photogrammetric Engineering and Remote Sensing*, vol. 62, no. 9, pp. 1067-1074, 1996.

[18] 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 Transactions on Geoscience and Remote Sensing*, vol. 40, no. 10, pp. 2300-2312, 2002.

[19] Z. Wang, D. Ziou, C. Armenakis, D. Li, and Q. Li, "A comparative analysis of image fusion methods," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 6, pp. 81-84, 2005.

[20] E. L. Pennec, and S. Mallat, "Sparse geometric image representations with bandelets," *IEEE Transactions on Image Processing*, vol. 14, pp. 423-438, 2005.

[21] D. Grois, I. Shcherback, T. Danov, and O. Yadid-Pecht, "Theoretical approach to CMOS APS PSF and MTF modeling - evaluation," *IEEE Sensors Journal*, vol. 6, pp. 118-124, 2006.

[22] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, "An MTF-based spectral distortion minimizing model for pan-sharpening of very high resolution multispectral images of urban areas," *In 2nd GRSS/ISPRS Joint Workshop on "Remote Sensing and Data Fusion over Urban Areas"*, IEEE Ed. (03EX646), 22-23 May 2003, Berlin, Germany, ISBN 0-7803-7719-2 : 90-94, 2002.

[23] A. Garzelli, and F. Nencini, "Interband structure modeling for Pan-sharpening of very high-resolution multispectral images," *Information Fusion*, vol. 6, pp. 213-224, 2005.

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

[25] M. Unser, "Splines: A perfect fit for signal and image processing," *IEEE Signal Processing Magazine*, vol. 16, no. 6, pp. 22-38, 1999.

[26] J.-M. Delvit, D. Léger, S. Roques, and C. Valorge, "Modulation transfer function and noise assessment," *In Proceedings of IGARSS 2003*, vol. 7, 4500-4502, 2003.

[27] D. Léger, F. Viallefont, P. Deliot and C. Valorge. *Post-Launch Calibration of Satellite Sensors*, On-orbit MTF assessment of satellite cameras, Morain & Budge Eds, Taylor and Francis Group, ISBN 90 5809 693 9, volume 2, pp. 67-75, 2004.

[28] C. Latry, B. Rougé, "Optimized sampling for CCD instruments: the Supermode scheme". *In Proceedings of IGARSS 2000 (Honolulu)*.

[29] R. Alonso Reyes, M.J. Gutierrez, S. Fernandez, C. Thomas, T; Ranchin, and L. Wald, "Evaluation of the quality of Quickbird fused products," *In Proceedings of the 24th Symposium of the European Association of Remote Sensing Laboratories (EARSeL)*, Dubrovnik, Croatia, May 25th-27th, 2004. *New Strategies for European Remote Sensing*, ed. by Marinko Oluc, Millpress Rotterdam Netherlands, p. 343-351, 2005.

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

[31] L. Alparone, S. Baronti, A. Garzelli, and F. Nencini, A global quality measurement of Pan-sharpened multispectral imagery. *IEEE Geoscience. Remote Sensing Letter*, vol. 1, no. 4, pp. 313-317, Oct. 2004.

[32] C. Thomas, and L. Wald, "A MTF-based distance for the assessment of geometrical quality of fused products". *In Proceedings of the 9th International Conference on Information Fusion, "Fusion 2006"*,

Florence, Italy, 10-13 July 2006, paper 267, IEEE Catalog 06EX1311C, 2006, ISBN 0-9721844-6-5 [CD-ROM]

[33] A. Papoulis, *Signal Analysis*, 3rd edition, McGraw-Hill, 1987.

[34] B. Aiazzi, B. Alparone, S. Baronti, and A. Garzelli, Context-driven fusion of high spatial and spectral resolution images based on oversampled multiresolution analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 40(10):2300-2312, 2002.

[35] C. Thomas, Fusion d'images de résolutions spatiales différentes. PhD Thesis, Ecole des Mines de Paris, Paris, France, 2006.

[36] M. M. Khan, L. Alparone, and J. Chanussot, "Pansharpening Quality Assessment Using the Modulation Transfer Functions of Instruments," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 11, pp. 3880-3891, 2009.



Pierre Massip was graduated engineering from the Ecole Spéciale de Mécanique et Electricité (Ivry-sur-Seine, France) in 2007 with a signal and image processing specialization. He obtained a Ph.D. degree from the MINES ParisTech in 2010. His research interests include image fusion in remote sensing imagery.



Philippe Blanc was graduated from the Ecole Nationale Supérieure des Télécommunications de Bretagne (France) in 1995 with a signal and image processing specialization. He obtained in the same year a M.S. degree in Signal, Telecommunications, Image and Radar from the University of Rennes (France) and a Ph.D. in Signal, Automatic and Robotics from the MINES ParisTech in 1999.

He worked in the Research Department of Alcatel Space, then Thales Alenia Space in projects related to Earth observation and spaceborne systems. Since 2007, he has been working as a senior scientist at the MINES ParisTech, Sophia Antipolis. He is the head of the research group "Observation, Modeling, Decision."



Lucien Wald received the M.S. degree in theoretical physics in Marseille and Paris, France, in 1977, the Ph.D. degree in Paris in 1980, and the Doctorat d'Etat ès Sciences on the applications of remote sensing to oceanography in Toulon, France, in 1985.

Since 1991, he has been a Professor at the MINES ParisTech, Sophia Antipolis, France. He is focusing his own research in applied mathematics and meteorology.

Prof. Wald is the recipient of the Autometrics Award in 1998 and the Erdas Award in 2001 for articles on data fusion. His career in information technologies has been rewarded in 1996 by the famous French Blondel Medal.