Image Degradation for Quality Assessment of Pan-Sharpening Methods

Wen Dou
Department of Geographic Information Engineering, Southeast University, Nanjing 210096, China; douw@seu.edu.cn
Received: 20 December 2017; Accepted: 20 January 2018; Published: 22 January 2018

Abstract: Wald’s protocol is the most widely accepted protocol to assess pan-sharpening algorithms. In particular, the synthesis property—which is usually validated on a reduced scale—is thought to be a necessary and sufficient condition of a success image fusion. Usually, the synthesis property is evaluated at a reduced resolution scale to take the original multispectral (MS) image as reference; thus, the image degradation method that is employed to produce reduced resolution images is crucial. In the past decade, the standard method has been to decimate the low-pass-filtered image where the filter is designed to match the modulation transfer function (MTF) of the sensor. The paper pointed out the deficiency of the method, and proposed a new image degradation method, referred to as method of spatial degradation for fusion validation (MSD4FV), which takes MTF compensation into account based on a simplified MTF model. The simulation results supported the implicit assumption of Wald’s protocol that image fusion performance is invariant among scales if the images have been properly degraded.

Keywords: remote sensing; image fusion; pan-sharpening; quality assessment; Wald’s protocol; MTF; scale; image degradation

1. Introduction

Currently, huge quantities of satellite images are available from many Earth observation platforms, and many of them have the ability to acquire panchromatic (PAN) and a multispectral (MS) images simultaneously. Because of signal-to-noise ratio (SNR) constraints and transmission bottleneck, MS images have a good spectral quality but a poor spatial resolution, whereas PAN images have a high spatial resolution but with a coarser/poorer spectral quality. Pan-sharpening is a branch of data fusion used to synthesize MS images at higher spatial resolution than original by exploiting the PAN high spatial resolution, which is important in the field of remote sensing, and many popular mapping products such as Google Maps/Earth use pan-sharpened imagery [1].

In the past three decades, many pan-sharpening methods have been developed [2–4]. Most algorithms fall into one of the two categories: component substitution (CS) and multiresolution analysis (MRA). CS methods enhance the spatial quality of MS by exploiting the spatial details of PAN, and MRA methods enhance the spectral quality of PAN with the spectral information of MS. All of these algorithms can be accommodated in a unique framework [5,6] which takes the fused image as a combination of a low-resolution MS image and modulated spatial details extracted from PAN.

Pioneering pan-sharpening algorithms [7] focused on visual enhancement and quantitative quality validation issues were at an early stage, while the guideline of second generation techniques is to meet the requirement of high-quality synthesis of spectral information [8], because it is very important for most remote sensing applications based on spectral signatures. That makes the quality assessment of pan-sharpened MS images a fundamental task. However, it is still a much debated topic.

As pan-sharpening techniques aim to construct high-resolution MS images that the corresponding sensor would observe, a straightforward solution to quality validation is to compare fused images...
with reference images, either by simulating the target sensor by means of high-resolution data from an airborne platform [9], or by degrading all available data to a coarser resolution and carrying out fusion from such data [10]. The quality assessment of fused image are taken as indicators of performance of the pan-sharpening algorithm, then different algorithms could be compared according to scores obtained from same test data.

Wald’s protocol [10] is the most widely accepted protocol, which implicitly assumes that the performances of the pan-sharpening methods are independent of the spatial resolution of the fused products; however, it is reported that the hypothesis is not always proper in practice [11]. To overcome this shortcoming, two strategies have been developed. The first strategy is to employ indices that do not require the availability of the reference image, such as Quality with No Reference (QNR) [12] or Khan’s protocol [13], to assess pan-sharpened MS images by some predetermined rules, which introduce more assumptions and problems. The second strategy which came into being recently inspects the trend instead of single measurement at a given degraded scale by Wald’s protocol [14,15]. However, these methods introduced more assumptions, and Wald’s protocol is still the most reliable.

Although there is doubt and dispute about the underlying assumption that the synthesis property is consistent among spatial scales, Wald’s protocol is still the most widely used protocol today, so it is of the highest priority to validate the assumption of Wald’s protocol. The paper illustrates that the assumption is good enough if images have been properly degraded. Section 2 introduces a new image degradation method, referred to as method of spatial degradation for fusion validation (MSD4FV), and Section 3 verifies MSD4FV with a real data set. Section 4 comes to three major conclusions: (1) Wald’s protocol is feasible; (2) Q2 n is presently the most suitable quality measurement; (3) MRA-based pan-sharpening algorithms will benefit from the proposed image degradation method.

2. Image Degradation Method for Wald’s Protocol

Wald et al. [10] defined the properties required in the fused image, and the protocol has been the most widely used one for validation of pan-sharpening methods. The first property is the consistency property, which requires that the original MS image can be obtained from the fused image by degradation. The second and third properties are usually referred to as the synthesis property, which requires that the fused image should retain the characteristics of the original MS image at a higher resolution.

To evaluate the synthesis property, one must compare the fused image with a reference high-resolution MS image, which is not always available. Assuming that the performance of fusion methods are invariant among scales, the problem could be solved by spatially degrading both the original MS and PAN images to coarser resolution so that the original MS image is used as a reference for the evaluation of the results [10]. Images are degraded by low-pass filtering and decimation according to scale ratio. However, the protocol has never mentioned which image degradation filter should be used. Aiazzi et al. fixed a degree of freedom in Wald’s protocol by specifying the filter for image degradation [16].

2.1. Aiazzi’s Method

Aiazzi et al. [16] stated that the spatial degradation method must take the modulation transfer function (MTF) of the imaging system into account, and proposed that MTF-tailored MRA methods optimized at coarser scale are still effective at finer scale. In [16], the low-pass filter for image degradation was designed to match the MTF of the sensor. Aiazzi’s method has become the standard implementation of image degradation for Wald’s protocol in the past decade. Recently, Vivone et al. contributed a MATLAB toolbox to the community which includes the procedure [17].

However, as [14] mentioned, fusion at degraded scale uses the cascade of the low-pass filter of the fusion methods with the lowpass filter used for decimation. This means fusion at full scale and degraded scale uses different filters. As the original MS is the degradation of an ideal image, the problem also exists in the process of image degradation. Figure 1 illustrates this effect by
constructing an image sequence using different strategies. The first strategy is to construct images with resolution ratio of 2, 4, 8, 16, and 32 times to original image by iteratively degrading the image by factor of 2, denoted as scale 1 to 5. The second strategy directly degrades the original image to these scales. Then, the images with the same scale are compared, and the root mean square error (RMSE) between images are plotted as function of scale. Aiazzi’s method—which is implemented by [17] as a standard implementation—shows remarkable inconsistency between the two strategies. When we degrade the image by a factor of 4 (which is the most common case in fusion validation), the RMSE of Aiazzi’s method was 27.520 (which is comparable to the standard deviation of the high-frequency component), while the RMSE of the method proposed in this paper and bicubic interpolation are 2.709 and 2.033, respectively. The observation implies that Aiazzi’s method for spatial degradation is problematic.

Figure 1. Aiazzi’s method is tested along with bicubic interpolation and the proposed method MSD4FV. The y-axis is the root mean square error (RMSE) between images constructed by direct and iterative spatial degradation strategies, and the x-axis is scale. The testing image is band 5 of a WorldView-2 image for the city of Beijing, China

2.2. MTF Compensation

Aiazzi’s method represents great progress towards precise quantitative research in fusion validation research. However, it ignores compensation before degradation with MTF-matched filter, which caused different equivalent filters at full and degraded scales—in other words, different MTFs among scales.

Another problem is that the decimation process is not carefully handled, which might lead to a shift of images. Although zero-phase filter and standard decimation would not introduce shift to digital signals, the situation is a little different in remotely sensed image processing. The difference is that element of an image is not measurement at a point but integral of measurement over an area. When decimation rate is even order, the decimation will introduce a shift of 0.5 high resolution pixel along x and y direction because the sample, i.e., pixel, in lower resolution image corresponds to a larger area, and the center could not coincide with the nearest higher resolution pixel.

The MTF of an image system can be modelled as [18,19]:

\[
\text{MTF} \quad \text{of} \quad \text{an} \quad \text{image} \quad \text{system} \quad \text{can} \quad \text{be} \quad \text{modelled} \quad \text{as} \quad [18,19]:
\]
MTF_{net}(x) = MTF_{opt}MTF_{det}MTF_{IM}MTF_{el}, \quad (1)

where \( x \) is spatial frequency, \( MTF_{opt} \) is the MTF caused by optics, \( MTF_{det} \) is the MTF caused by the detector, \( MTF_{IM} \) is the MTF caused by image motion, and \( MTF_{el} \) is the MTF caused by electronics. As \( MTF_{el} \) is close to 1 and \( MTF_{IM} \) at the cross-track direction is 1, Equation (1) can be approximated as

MTF_{net}(x) = MTF_{opt}MTF_{det} . \quad (2)

Usually, \( MTF_{opt} \) is modelled as a Gaussian function:

MTF_{opt}(x) = e^{-x^2/2\sigma^2}, \quad (3)

and \( MTF_{det} \) is modelled as [18]

MTF_{det}(x) = \text{sinc}(0.5\pi x/x_{Nyq}), \quad (4)

where \( \text{sinc}(x) = \sin(x)/x \). As the nominal MTF value of a sensor is the gain at Nyquist frequency \( x_{Nyq} \), and \( MTF_{det} \) at the Nyquist frequency is \( MTF_{det}(x_{Nyq}) = \text{sinc}(0.5\pi) \), \( MTF_{opt} \) at the Nyquist frequency is

MTF_{opt}(x_{Nyq}) = \frac{MTF_{net}(x_{Nyq})}{MTF_{det}(x_{Nyq})} = \frac{MTF_{net}(x_{Nyq})}{\text{sinc}(0.5\pi)}. \quad (5)

From Equations (3) and (5),

\[ \sigma = \left[ \frac{x_{Nyq}^2}{2(\ln \text{sinc}(0.5\pi) - \ln MTF_{net}(x_{Nyq}))} \right]^{1/2}, \quad (6) \]

so \( MTF_{opt} \) is determined. By separating MTF into \( MTF_{opt} \) and \( MTF_{det} \), we can compensate these two components separately according to corresponding models. This is crucial because \( MTF_{det} \) is greatly related to sampling, which is ignored in previous research.

2.3. Method of Spatial Degradation for Fusion Validation

To amend Aiazzi’s method, the paper proposes a new method of spatial degradation for fusion validation, referred to as method of spatial degradation for fusion validation (MSD4FV). The main idea is that MTF compensation should be conducted before image degradation, and \( MTF_{det} \) and \( MTF_{opt} \) are processed separately according to a simplified MTF model in Equation (2). The basic steps are as follows:

1. Fourier transform (FT) the image;
2. Divide the FT of the image by \( MTF_{det} \) (Equation (4)) to compensate \( MTF_{det} \);
3. Divide the result of the previous step by \( MTF_{opt} \) (Equation (3)) to compensate \( MTF_{opt} \);
4. Degrade according to \( MTF_{opt} \) with cutoff frequency at \( 1/r \) of the Nyquist frequency of the original image, where \( r \) is the scale ratio;
5. Inverse Fourier transform (IFT);
6. Resize the image by means of average with ratio of \( 1/r \).

It is notable that the last step fulfills the degradation of \( MTF_{det} \) and decimation simultaneously. Although the whole process could be conducted in spatial domain by designing spatial filters based on MTFs, MSD4FV is presented in the current form to help readers understand and focus on the idea.
3. Experiments and Results

3.1. Data Set

A WorldView-2 image subset acquired over the urban area of Beijing, China is used because of an abundance of spatial details. The data sets are composed of a PAN image of 8192 × 8192 pixels, eight MS image bands consisting of 2048 × 2048 pixels, and the scale ratio between PAN and MS bands is 4.

To test the influence of spatial degradation on fusion validation, a sequence of PAN/MS images with decreasing spatial resolutions was constructed by iteratively degrading the original images by a factor of 2. The sequence of PAN/MS images is denoted as \( \{\text{PAN}_i/\text{MS}_i\} \), where PAN\(_0\)/MS\(_0\) are original images, and \( i \) is the relative scale to original images. The spatial resolution ratio between PAN\(_i\) and MS\(_i\) was 4, and that between PAN\(_j\)/MS\(_j\) and PAN\(_k\)/MS\(_k\) was \( 2^{j-k} \). All couples of PAN\(_i\)/MS\(_i\) are fused and then compared to MS\(_{i-2}\), so only couples with \( i \geq 2 \) are fused. PAN images were degraded by bicubic interpolation since PAN images have been MTF compensated by vendors [16], and MS images were degraded by Aiazzi’s method and MSD4FV, respectively.

3.2. Pan-Sharpening Algorithms and Quality Indices

Vivone et al. has contributed a MATLAB toolbox to the community [17] which implements the most widely used pan-sharpening procedures and quality indices. All of the 18 pan-sharpening algorithms compared in [17] are chosen to make the research comparable. Quality indices include \( Q^2_n \) [20], Relative dimensionless global error in synthesis (ERGAS) [21], and the Spatial Correlation Coefficient (SCC) [22]. \( Q^2_n \) and SCC evaluate how two images are similar, and two identical images scores 1. ERGAS evaluates how two images are different, so the ideal score is 0.

3.3. Results

Two sequences of PAN/MS images were constructed by Aiazzi’s method and MSD4FV, respectively. For each sequence, all couples of PAN\(_i\)/MS\(_i\) that \( i \geq 2 \) were fused, and the selected quality indices were assessed with MS\(_{i-2}\) as reference. For each quality index, the scores were plotted against scales at which pan-sharpening procedures were carried out.

Figure 2 reports \( Q^2_n \) scores as a function of the spatial scale of the PAN/MS images to be fused. The image sequence constructed following Aiazzi’s method illustrates a remarkable dependence from the spatial scale, decreasing with the spatial scale. Meanwhile, the sequence of the proposed method illustrates much less dependency on spatial scale. For all pan-sharpening algorithms, the proposed degradation method scored higher than Aiazzi’s method at any scale. Particularly, the difference at scale 2 is of most interest because that is the scale where quality assessment is usually taken. A significant difference can be observed for several algorithms, such as BDSD, Indusion, ATWT M3, which means that these algorithms might have been underestimated in previous researches. For example, using Aiazzi’s method, IHS is a bit better than Indusion in terms of \( Q^2_n \), while Indusion performs much better than IHS with the proposed method.

Figures 3 and 4 report ERGAS and SCC scores as a function of the spatial scale of the PAN/MS images to be fused, respectively. These two results also illustrate scores closer to ideal for MSD4FV. For the ERGAS result (Figure 3), MSD4FV illustrates much better consistency among scales—especially adaptive algorithms such as BDSD, GSA, and PRACS. The SCC result (Figure 4) of the proposed method implies nearly complete independence of scale for most pan-sharpening algorithms, which means that the implicit assumption of Wald’s protocol is most likely reasonable.
Figure 2. Q2n scores of fused image sequences constructed following Aiazzi’s method and MSD4FV.

Figure 3. Relative dimensionless global error in synthesis (ERGAS) scores of fused image sequences constructed following Aiazzi’s method and MSD4FV.
As an example, Figure 5 compared performance of GS and Indusion using different image degradation methods, measured by $Q_2^n$. When images are degraded following Aiazzi’s method (dotted lines), GS scores much higher than Indusion, while assessment on degraded images using MSD4FV (solid lines) shows that Indusion is a little better than GS. Figure 6 illustrates how image degradation influences assessment of pan-sharpening methods. The first row shows a small portion of original MS image, along with images fused by GS (b) and Indusion (c) at full resolution scale. Both algorithms perform similar in spatial enhancement and GS is a little better, meanwhile Indusion has much higher spectral fidelity. The second row shows MS image degraded by Aiazzi’s method and the pan-sharpened images at reduced resolution scale, and the last row shows MS image degraded by MSD4FV and the pan-sharpened images. Visual check of (e) and (f) might lead to the conclusion that GS is a much better algorithm than Indusion, while (h) and (i) hold the conclusion drawn from fused result at full resolution, which is what we have expected for Wald’s protocol.
4. Discussion and Conclusions

As Aiazzi’s method would introduce inconsistency among scales, as Figure 1 implies, this paper proposed an improved spatial degradation method for Wald’s protocol. Simulation shows that when images are spatially degraded by MSD4FV, the performance of pan-sharpening algorithms manifest weak dependence on spatial scale, which supports the hypothesis assumed by Wald’s protocol.

In particular, the SCC quality index illustrates the least dependence; however, the difference among pan-sharpening algorithms is also the least, which means that SCC might not be good to discriminate. There is no linear trend for ERGAS that could be observed for SCC and Q2^n, and most algorithms reach their peaks at scale of 4 or 5. A possible explanation is that such scales are related to the size of buildings and roads, thus the ERGAS is more sensitive to landscape than Q2^n and SCC, which needs more investigation. Although Q2^n is monotonically decreasing with scale for either
method, MSD4FV yields a much narrower range. So, Q2 is the most suitable single measurement for quality assessment among the tested indices.

An interesting observation is that adaptive pan-sharpening methods such as BDSD, GSA, MTF-GLP-CBD show minimum dependence on scale. As so-called adaptive algorithms should theoretically perform consistently over scales or landscape patterns, this observation validated the proposed spatial degradation method and these adaptive pan-sharpening algorithms simultaneously.

This paper is a preliminary work, and there are still many open points. First of all, more data sets with different sensors over different landscapes must be tested, and more quality indices should be considered. How the accuracy of nominal MTF would influence quality assessment should be studied, which might lead to the question of how to validate pan-sharpening when MTF is unavailable or inaccurate. Techniques used by adaptive methods mentioned in the letter, along with others handling information among scales such as SIFT and Kalman Filter, might help in the study. Consistency property measurement is another crucial problem concerning spatial degradation. Study on how the degradation method influence scores of quality indices for a given fusion algorithm would help us understand image fusion more deeply. Finally, the proposed degradation method is actually a new method to extract spatial details, so most MRA-based pan-sharpening algorithms would benefit from the idea of MSD4FV.

Acknowledgments: The author is grateful to the anonymous reviewers, whose insightful comments have notably improved the clarity of presentation of the paper. This research was funded by the National Natural Science Foundation of China (no. 41401382, no. 41571342) and China Geological Survey Project (no. DD20160076)

Conflicts of Interest: The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References


