A CRITICAL REVIEW OF QUALITY ASSESSMENT PROTOCOLS IN PAN-SHARPENING

Shiva Aghapour Maleki 1, Hassan Ghassemi 1,*

1 Image Processing and Information Analysis Laboratory, Faculty of Electrical and Computer Engineering, Tarbiat Modares University, Tehran, Iran

s_aghapourmaleki@modares.ac.ir, ghassemi@modares.ac.ir

Commission VI, WG VI/4

KEY WORDS: Quantitative Assessment, Pan-Sharpening, Image Fusion, Multispectral Image, Panchromatic Image

ABSTRACT:

Although vast amounts of pan-sharpening methods have been proposed to date, there has been relatively little published on the topic of qualitative and quantitative assessment of the pan-sharpened multispectral (MS) data. Since a high resolution reference MS image is not available, qualitative and quantitative assessments of the spatially enhanced MS image are a much more challenging task in pan-sharpening. Thomas et al. conducted a critical survey of conventional pan-sharpening methods considering remote sensing physics. In this paper, we study the effects of physical constraints of satellite sensors on conventional quality assessment protocols. The novelty of this work is in investigating the effect of physical constraints on the performance of quality assessment protocols. The most popular protocols, which are analysed here, are quality not requiring a reference (QNR), Wald’s protocol, Zhou’s protocol, Khan’s protocol and Ghassemi’s protocol. In doing so, the strengths and weaknesses of such protocols regarding the physical limitations on satellite sensors are discussed.

1. INTRODUCTION

Quality assessment of pan-sharpened multispectral (MS) images is a critical task. The conventional quality assessment protocols were designed to measure the quality of the pan-sharpened results while the MS and panchromatic (PAN) data are ideal, i.e., there is no dissimilarity between them. However, physical and technical constraints on designing satellite sensors do not allow us to provide data with coupling spectral and spatial resolutions. As will be discussed later, the sensors may provide the PAN and MS data that exhibit dissimilarities such as object disappearances, contrast inversion or contrast reversal information, and an unknown number of dissimilarities between the PAN and MS data. These limitations enforce us to consider them while designing a quality assessment protocol. In fact, a well-established quality assessment protocol should take such constraints into account (Gahremani and Ghassemian, 2016), (Ghassemian, 2000), (Li, 2000), (Thomas and Wald, 2004), (Thomas et al., 2008), (Vivone et al., 2018).

So far, various quantitative assessment metrics have been proposed to assess the fused MS products. Some quantitative indexes have been developed for pan-sharpening such as Q4 (Alparone, et al. 2004), relative dimensionless global error in synthesis (ERGAS) (Wald, 2000), and spatial correlation coefficient (SCC) (Zhou et al. 1998), and several others have been developed for image processing which is also utilized in pan-sharpening like universal image quality index (UIQL, which is denoted by Q in equations) (Wang and Bavik, 2002), spectral angle mapper (SAM) (Yuhas et al. 1992), correlation coefficient (CC), and root mean square error (RMSE) (Ranchin and Wald, 2000), (Wald, 2000). Due to the fact that high resolution MS data are unavailable, we first need to define a formal framework before applying a quantitative index to the pan-sharpened results. The word “quality assessment protocol” or briefly “protocol” is used to refer a formal framework. Thus far, several protocols have been presented in the literature. Wald’s protocol (Ranchin et al. 2003), (Wald, 2000), Zhou’s protocol (Zhou et al. 1998), quality not requiring a reference (QNR) protocol (Alparone et al. 2008), Khan’s protocol (Khan et al. 2009), and Ghassemi’s protocol (Ghassemian, 2016) are generally utilized in pan-sharpening. Of the five mentioned protocols, Wald’s protocol is the most common. In the design procedure of such protocols, it is assumed that the MS and PAN data are well-behaved (in this study, well-behaved data refer to the MS and PAN data that have complementary information without any dissimilarity between them). In (Thomas et al. 2008) published their experiments on the characteristics of satellite data and their influence on pan-sharpening methods. This research shows that the available MS and PAN images have a number of dissimilarities that affect the pan-sharpening methods. The essential question is whether the conventional protocols take these limitations into account.

This survey and the emphasis within this paper on such protocols are based on the characteristics of the MS and PAN data in reality. In fact, while a pan-sharpening method is applied to real MS and PAN data, we need to use a protocol that is suitable for real data. Establishing a suitable protocol can better show the strengths and weaknesses of a pan-sharpening method. We will show that the conventional protocols are unable to measure the fusion results in challenging regions (in this study, challenging regions refer to the MS and PAN data that have different dissimilarities). Since there are different methods for pan-sharpening, in this study we use PCA and ATWT methods that are from two different categories of pan-sharpening.
methods, called Projection–Substitution Methods and Multiscale Models respectively.

The rest of the paper is structured as follows. Section 2 briefly reviews the aforementioned protocols. In Section 3, we analyze the effects of dissimilarities between the PAN and MS images on the quality assessments protocols. Discussion and conclusions are provided in Section 4.

2. QUALITY ASSESSMENT PROTOCOLS

2.1 Notations

Let $M$ and $P$ be the MS and PAN images at their original spatial resolutions, respectively. Subscripts $d$ and $r$ indicate the down-sampled and the resampled versions of an image. The pan-sharpened MS image is denoted by $\hat{M}$ and the average of the fused image on its spectral bands is denoted by $\hat{P}$. In this paper, $L$ is the number of spectral bands, and HSR is abbreviation of high spatial resolution.

2.2 Wald’s Protocol

Wald’s protocol is a well-established protocol, which relies on three properties. According to this protocol, a pan-sharpening method should satisfy the following three requirements (Ranchin and Wald, 2000), (Ranchin et al. 2003), (Thomas and Wald, 2006), (Wald, 2006):

- Consistency property: any synthetic image $\hat{M}$ once degraded to its original resolution, should be as identical as possible to the original image $M$. 
- Synthesis property (the first part): any image $\hat{M}$ fused by means of an HSR image should be as identical as possible to the ideal $M$ image that the corresponding sensor, if existent, would observe at the resolution of the HSR image.
- Synthesis property (the second part): the multispectral vector of $\hat{M}$ images fused by means of an HSR image should be as identical as possible to that of ideal $M$ images that the corresponding sensor, if existent, would observe at the spatial resolution of the HSR image.

Due to lack of HSR MS images, the synthesis properties cannot be directly verified. To deal with this problem, synthesis is usually checked at degraded spatial scales. For this purpose the PAN and MS images are filtered by a proper low-pass filter and then decimated by the factor of scale ratio between them, then a pan-sharpening method is applied on the degraded images. The resulting fused image is at the resolution of original MS image and therefore, the original MS image can play the role of the reference image and synthesis property can be checked.

2.3 Zhou’s Protocol

Zhou’s protocol assesses the spectral and spatial qualities separately at full resolution without any degradation. According to this protocol (Zhou et al. 1998), the spectral quality of the fused image is measured for each band by the average absolute differences between the fused image $\hat{M}$ and corresponding resampled MS image $M_r$. The obtained value should be as close to zero as possible. To calculate the spatial quality, first spatial details are extracted from the PAN and MS images through Laplacian filter, and then the spatial correlation coefficient (SCC) is calculated between the extracted high-pass information from the PAN and each band of the pan-sharpened MS images. The closeness of the resulting value to 1 indicates the spatial similarity of two images.

2.4 Quality with No Reference (QNR) Protocol

The quality with no reference protocol (QNR) also measures the quality of fused images at full resolution. The difference of this protocol to other mentioned protocols is that it does not require a reference image. This protocol measures spectral and spatial distortions that are shown as $D_{4,q}^{\text{QNR}}$ and $D_{4,r}^{\text{QNR}}$ respectively. These two distortion metrics, which use the UIQI index to determine the degree of distortion, are defined as:

$$D_{4,q}^{\text{QNR}} = \frac{1}{L(L-1)} \sum_{i=1}^{L} \sum_{j=1}^{L} |Q(M^{(i)}, M^{(j)}) - Q(\hat{M}^{(i)}, \hat{M}^{(j)})|$$

and

$$D_{4,r}^{\text{QNR}} = \frac{1}{L} \sum_{i=1}^{L} |Q(M^{(i)}, P) - Q(\hat{M}^{(i)}, P)|$$

In these equations, $p$ and $q$ are positive integers which chosen to emphasize large difference values. They are set equal to one by default. $D_{4,q}^{\text{QNR}}$ indicates that the inter-relation between the fused bands should be as identical as the relation between MS bands and $D_{4,r}^{\text{QNR}}$ indicates that the relation between each fused band and PAN image should be identical to those between MS band and low resolution PAN. By combining these two distortions a unique quality index named QNR can be obtained (Alparone et al. 2008), (Carlà et al. 2015).

2.5 Khan’s Protocol

Similar to the two former protocols, Khan’s protocol also comprises two indexes, one for spatial distortion and the other for spectral distortion (Khan et al. 2009). The metric used to measure the degree of distortion is Q4. The foundation of this protocol is based on Wald’s consistency property. This protocol uses the basis of Zhou’s protocol (extracting high-pass information) to measure the spatial distortion and basis of QNR protocol to measure spectral distortion (Carlà et al. 2015). In fact, Khan et al. tried to improve the previous protocols by using the strengths of each of the aforementioned protocols. To quantify the spectral distortion at first modulation transfer function (MTF) filters are applied to the fused MS bands and then they are decimated, then the Q4 index is calculated between each band of the obtained image and its low resolution original, afterward an average over the obtained Q4 values is taken and finally by subtracting the average by 1 the final value is obtained:

$$D_{4}^{\text{Khan}} = 1 - Q4(\hat{M}, M)$$

To define the spatial distortion, the authors in (Khan et al. 2009) used the high-pass complements of the MTF filters to extract the high-frequency information from images. The process is as follows: at first MTF filter is applied to each band of fused MS image, then the filtered image is subtracted from
the corresponding fused MS image to obtain the details of fused image. This procedure is also applied to original PAN image to obtain its details. The Q4 index is calculated between the details of fused image and PAN image. The same method is applied on original MS and low resolution PAN images. Finally by calculating the absolute difference in the Q4 values across scales the spatial distortion is obtained:

\[
D_{Khan}^4 = \frac{1}{L} \sum_{i=1}^{L} |Q(M^h, P_d) - Q(\hat{M}^h, P_d)|
\]  

(4)

where the superscript \( h \) denotes highpass information.

2.6 Ghassemi’s Protocol

Ghassemi’s protocol also assesses the spectral and spatial qualities separately (Ghassemian, 2016). To assess the spatial quality, by taking a weighted average over the bands of the fused image, a PAN image \( \hat{P} \) is obtained. This image is compared with the original PAN image \( P \); the used comparison method can be any distance measure that can calculate the spatial differences between two single PAN images, for example ERGAS. To calculate the spectral quality, the fused image \( \hat{M} \) is down-sampled to produce the low version of original MS. The result is called \( \hat{M}_d \). Then the dissimilarity between \( \hat{M}_d \) and \( M \) is calculated. The used dissimilarity measure can be any distance measure that can calculate the spectral differences between two multispectral images, for example SAM.

3. INVESTIGATION OF REMOTE SENSING PHYSICS ON QUALITY ASSESSMENT PROTOCOLS

In this section, the robustness of the quality assessment protocols briefly reviewed in the previous section are appraised by several dissimilarities existed in real data.

3.1 Object Disappearances

Object Disappearances are commonly observed in remotely sensed data. This phenomenon is chiefly occurred due to the different spectral bands of acquisition sensors of the PAN and MS data (Thomas, 2008). Figure 1 shows an example of object disappearances. Figure 1(a)-(d) shows a region of the image acquired in four bands, namely, red, green, blue, and near infrared (NIR), respectively. The corresponding PAN excerpt degraded to the original spatial resolution of the MS image is presented in Figure 1 (e). The edges are clearly visible in the PAN image are missing in the red image and are weak in the green and the blue images.

In the following, we investigate the performance of the mentioned five protocols in these regions. To evaluate them, we assume only the red image has missing information. If Wald’s and Ghassemi’s protocols, the spectral distortion metrics of Zhou’s and Khan’s protocols are utilized to assess the fusion results, the red band of the original MS image plays as a reference to the corresponding band of the fused image. According to these protocols, the added spatial details to the red band of the fused results are considered as spectral distortion because there is a large differences between the fused and the reference images in this band. In other words, the reference image considers the missing information as spectral distortion while this is not true. The \( D_{SNR}^4 \) metric uses inter-band UIQI values. Thus, we assume that our MS dataset has the red and the NIR images. Owing to the fact that we inject common information, i.e., spatial details, to the red and the NIR images, the UIQI between the mentioned bands will be increased after fusion; thus, the UIQI value after fusion is larger than the UIQI value before fusion. From Eq. (1), any difference in the UIQI values before and after fusion is assumed as spectral distortion; hence, according to this formula we have a significant distortion.

Figure 1. Illustration of object disappearances in satellite imagery (a) Red band. (b) Green band (c) Blue band (d) NIR band. (e) PAN band down-sampled to spatial resolution of MS.
in the fused red and the fused NIR bands while these bands are successfully improved spatially after fusion.

Now we evaluate the spatial distortion metrics of the QNR, Zhou’s and Khan’s protocols for the red band of the MS data.

- $D_{\text{Khan}}$: The UIQI after fusion is bigger than the UIQI value before fusion (This is because the similarity between the PAN image and the fused red band is much more than the similarity between the degraded PAN image and the red band of the MS image). This means that we have spatial distortion in the pan-sharpened red band through this index.

- $D_{\text{QNR}}$: Before pan-sharpening, the relationship between the red band of the MS image and the down-sampled PAN image is weak; therefore, the UIQI between them is small. On the other hand, the UIQI between the fused red band and the PAN image is considerably increased after pan-sharpening. According to this protocol, the difference between UIQI values before and after fusion indicates that there exists spatial distortion in the fused red band.

- Zhou’s spatial distortion metric: As explained before, this metric uses the CC to measure the similarity between the high frequency information of the MS image and that of the PAN image. Because the metric uses the PAN image as a reference, the results of this metric are more reasonable than those of the other three protocols. However, the main drawback of this protocol according to (Wang, 2002), is that the CC itself cannot correctly calculate the similarity of two images.

The numerical results of above discussions are presented in table 1:

<table>
<thead>
<tr>
<th>Quality Index</th>
<th>Before Fusion</th>
<th>After Fusion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PCA</td>
<td>ATWT</td>
</tr>
<tr>
<td>UIQI ($D_{\text{QNR}}$)</td>
<td>0.53</td>
<td>0.76</td>
</tr>
<tr>
<td>UIQI ($D_{\text{Khan}}$)</td>
<td>0.36</td>
<td>0.70</td>
</tr>
<tr>
<td>UIQI ($D_{\text{QNR}}$)</td>
<td>0.73</td>
<td>0.98</td>
</tr>
</tbody>
</table>

Table 1. Numerical results of assessment for object disappearances

3.2 Contrast Inversion

Contrast inversion is another category of dissimilarities that commonly occurred in the satellite locally or globally. Figure 2 shows an example of contrast inversion. In this case, the NIR image as well as the PAN image is in inverse pattern with the rest bands of the MS image. In the following, we evaluate the impacts of the contrast reversal information on the performances of the protocols when assessing the fused images.

1. QNR:

- $D_{\text{QNR}}$: As pointed out before, to evaluate this metric, we need two spectral bands at least. We assume our data set contains the red and the NIR bands of the MS image as well as the corresponding PAN image, in which the MS bands are in contrast inversion. Before pan-sharpening, the CC between the red and the NIR images is a small value.

After pan-sharpening, common spatial details are injected into the MS bands, and thus the measured CC between the pan-sharpened red and the pan-sharpened NIR images gets larger than the CC calculated before pan-sharpening.
Hence the UIQI value becomes larger than the UIQI before fusion, and this means the spatial details introduces spectral distortion in the pan-sharpened bands.

- $D^{QNR}$: Similar results hold for the spatial quality index when using Eq. (2). When the red and the PAN images are in contrast inversion, the UIQI value after fusion is larger than the UIQI value before fusion, and the difference between UIQI values is considered as spatial distortion.

- $D^{Khan}$: The CC between the red and the fused red images is a small value, and this makes the overall hypercomplex CC smaller; hence, contrast inversion results in significant decrease in the Q4 metric, and consequently, this metric shows that there is spectral distortion in fused results.

- $D^{zhou}$: Likewise, the term $Q(M^{h,0}, P^{h})$ is greater than the term $Q(M^{h,0}, P^{h})$, and according to this metric, spatial distortion is inevitable in the MS bands that show contrast reversal information.

3. Zhou’s protocol:

- Zhou’s spectral distortion metric: Because the fused red image contains the spatial details of the PAN image, which itself is in contrast inversion with the red image, the difference between the resampled red image and the fused red image is large, and we have a large distortion in the pan-sharpened red image using this metric.

- Zhou’s spatial distortion metric: This can measure the spatial distortion correctly if the contrast reversal information is occurred globally. As discussed before, this protocol cannot take local dissimilarities into account.

4. Ghassemi’s protocol:

When two images are in contrast inversion, their edges are also in contrast inversion. If the PAN’s spatial details are inversely injected into the red image, then the difference between the fused red image and the reference red image is small. Otherwise, the edges of the fused red image and the reference red image are in contrast inversion, and the injected information introduce both spectral and spatial distortions based on this protocol.

The numerical results of this section are as follows:

<table>
<thead>
<tr>
<th>Quality Index</th>
<th>Before Fusion</th>
<th>After Fusion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PCA</td>
<td>ATWT</td>
</tr>
<tr>
<td>$D^{QNR}$</td>
<td>0.23</td>
<td>0.51</td>
</tr>
<tr>
<td>$D^{QNR}$</td>
<td>0.59</td>
<td>0.87</td>
</tr>
<tr>
<td>$Q4$</td>
<td>-</td>
<td>0.29</td>
</tr>
<tr>
<td>$UIQI(D^{Khan})$</td>
<td>0.38</td>
<td>0.53</td>
</tr>
</tbody>
</table>

Table 2. Numerical results of assessment for contrast inversion

4. DISCUSSION AND CONCLUSION

In this paper, we critically review the performances of the conventional quality assessment protocols in pan-sharpening. For well-behaved data, the conventional quality assessment protocols, i.e., QNR, Wald’s, Khan’s, Zhou’s, and Ghassemi’s protocols, are suitable to measure the quality of pan-sharpened results. However, the remote sensing physics do not allow us to have a perfect sensor, and care must be taken to use these protocols in challenging regions. Indeed, we require a perfect protocol to calculate the quality of fused images in such regions. Otherwise, the dissimilarity characteristics between the PAN and MS data can affect the quality measurement of the fused results. The sophistication of designing a quality assessment protocol depends on the characteristics of the PAN and the MS images to be fused. Two types of the dissimilarities, namely, object disappearance and contrast inversion, have been studied here. In the former case, it was seen that the quality of the fused bands were quantitatively worse while the fused bands were spatially enhanced. Similar behaviour has been observed for quantitatively assessing the data with contrast reversal information. It is often written that there is trade-off between preserving spectral information and enhancing the spatial information (Alparone et al. 2008), (Batur and Maktav, 2019), (Ghassemian, 2001), (Khan et al. 2009), (Lillo-Saavedra and Gonzalo, 2006), (Marcello et al. 2013), (Valizadeh and Ghassemian, 2012) (Zhang et al. 2012), (Zhou et al. 1998) (Shahdoosti and Ghassemian, 2012). Bearing in mind that the characteristic of spatial details is spectral information with strong variations in topography; hence, this approach requires that the spatial information of the PAN data should be considered as complementary information of the MS bands, and the trade-off between spatial and spectral information is not fatality (Du et al. 2007), (Ghahremani and Ghassemian, 2015), (Thomas et al. 2008). Forthcoming studies concern the impacts of other dissimilarities on quality assessment protocols as well as on pan-sharpening methods, with the aim of designing a suitable quality assessment protocol, which is closer to real data.

REFERENCES


Wald, L., 2000, January. Quality of high resolution synthesised images: Is there a simple criterion?. In Third conference ‘Fusion of Earth data: merging point measurements, raster maps and remotely sensed images” (pp. 99-103). SEE/URISCA.

