JOINT QUALITY MEASURE FOR ACCURACY ASSESSMENT OF PANSHARPENING METHODS

Gintautas Palubinskas

German Aerospace Center DLR, Remote Sensing Technology Institute
Oberpfaffenhofen, 82234 Wessling, Germany

ABSTRACT

A new joint quality measure JQM, which is a sole measure is proposed for quality ranking of pansharpening methods. It is based on a newly proposed composite similarity measure CMSC, which consists of means, standard deviations and correlation coefficient and is translation invariant with respect to all parameters. JQM itself consists of a weighted sum of two terms. First term is measured between a low pass filtered pansharpened image and original multispectral image in a reduced resolution scale. The second one – between weighted intensity calculated from pansharpened image and original panchromatic image in a high resolution scale. Experimental results show advantages of a new measure JQM for quality assessment of pansharpening methods on the one hand and drawbacks of already known measure QNR on the other hand.

Index Terms— Multi-resolution, multi-sensor, image fusion, pansharpening, quality assessment measure

1. INTRODUCTION

Pansharpening aims to include spatial/detail information from a high resolution image (e.g. panchromatic/ multispectral image) into a low resolution image (e.g. multi-spectral/hyper-spectral image) while preserving spectral properties of a low resolution image. A large number of algorithms and methods were introduced to solve this problem during the last two decades which can be divided into two main groups. First group of methods is based on a linear spectral transformation (e.g. Intensity-Hue-Saturation IHS, Principal Component Analysis, Gram-Schmidt orthogonalization GS) followed by a Component Substitution (CS). Methods of the second group use spatial frequency decomposition usually performed by means of high pass filtering, e.g. boxcar filter in signal domain, filtering in Fourier domain or multi-resolution analysis MRA using wavelet transform. Here we have to mention that there are some attempts to combine both types of methods. Moreover, there exist a group of methods which state pansharpening task as an ill-posed recovery problem solved by regularization using Bayesian estimation and recently proposed sparse representation approaches. For recent surveys of various image fusion methods see references [1, 2].

In parallel to pansharpening methods development many attempts were undertaken to assess quantitatively their quality usually using measures originating from image/signal processing such as mean squared error (MSE), peak signal-to-noise ratio (PSNR), relative dimensionless global error in synthesis (ERGAS), Pearson's correlation coefficient (CC), spectral angle mapper (SAM), universal image quality indices (UIQI/SSIM) and multispectral extensions of UIQI ($Q4/Q2^n$). For recent overview of quality measures see references [3, 4, 5]. These simple/separate measures defined in scalar/vector form can be used only as full reference measures. Due to the missing reference in pansharpening quality assessment task different solutions or so called protocols were proposed: Wald's, Zhou's, Quality with No Reference (QNR) and Khan's [3], which usually include calculation of several quality measures. Of course a sole or joint quality measure as already proposed in [6-8] enables easier und much more practical/comfortable ranking of various fusion methods.

Usually pansharpening in image processing is used to increase visual quality of an image. In remote sensing this task is fully different because it aims at enhancing image quality for further processing such as clustering, classification, matching and change detection thus requiring only relative comparison of data (translation invariant applications). Recently spreading quality measure UIQI/SSIM [9] was designed for perceptual tasks (scale invariant applications) thus its usage in pansharpening quality assessment in remote sensing imagery (see e.g. QNR [6] and joint quality measures [7, 8]) can lead to wrong results. Because MSE and UIQI/SSIM based measures are not well suitable for translation invariant (with respect to sample means and standard deviations) applications [10] we propose to exchange/replace above mentioned measures by a new measure – Composite measure based on Means, Standard deviations and Correlation coefficient CMSC [10], which is translation invariant with respect to all parameters, thus enhancing measures proposed in [8].

In this paper we perform a comparison of six pansharpening methods originating from the main groups of

Proc. IGARSS, Milan, Italy, IEEE, pp. 601-604, 2015

methods using a new joint quality measure JQM and already known measure QNR for IKONOS satellite data.

2. PANSHARPENING METHODS

Methods investigated in this paper can be described by the following general expression (see e.g. [11, 12, 1])

$$msf_k = msi_k + g_k \cdot (pan - pan_{lpf}), \tag{1}$$

where msf_k – fused high resolution multispectral image, k – spectral band number, msi_k – interpolated low resolution multispectral image, g_k – weight (gain) for detail injection, pan – high resolution panchromatic image and pan_{lpf} – low pass filtered pan image. Then individual methods can be seen as special cases of (1) as shown below. Thus General Fusion Filtering (GFF) [13] is defined as $msf_k = FFT^{-1}\left[\left(ZP\left(W \cdot FFT(ms_k)\right) + FFT(pan) \cdot (1 - LPF)\right)\right]$, where $g_k = 1$, ms_k – low resolution multispectral image, ZP – zero padding, W- Hamming window and LPF – low pass filter. High Pass Filtering Method HPFM (variant of GFF) [14] is given by (1) with $g_k = 1,\ pan_{lpf} = pan * lpf,\ lpf = FFT^{-1}(LPF)$. Ehlers fusion [15] is defined as $msf_k = msi_k - I + I * lpf1 + pan - pan * lpf2$, where $g_k = 1$, intensity is defined as $I = \sum w_k \cdot msi_k$ and w_k are spectral weights. Two different low pass filters are used for filtering of pan and intensity images respectively (Paper author's implementation is used). Á trous wavelet transform ATWT [16] is given by (1) with $g_k = 1$ and pan_{lpf} - à trous wavelet decomposed low resolution version of pan (M. Canty's implementation [17] is used). Component substitution using IHS transformation (CS IHS) is (1) with $g_k = 1,\ pan_{lpf} = I$ (Paper author's implementation is used). Component substitution using GS transformation (CS GS) is (1) with $pan_{lpf} = I$ (IDL ENVI implementation is used).

3. QUALITY ASSESSMENT MEASURES

Quality or similarity measures can be divided into two main groups: full reference (FR) measures when the reference image is existent (quite few applications mostly simulations) and no reference measures (most applications). Examples of full reference measures used to assess pansharpening quality are SAM, MSE based e.g. PSNR and ERGAS, CC, UIQI/SSIM and Q4 just to mention few of them. It was shown in [10] that MSE based measures are not translation invariant with respect to sample standard deviation and recently widely spreading SSIM measure is not translation invariant with respect to both sample moments: mean and standard deviation. This can lead to wrong quality assessment results in applications requiring translation invariance property (only relative comparison of parameters is required independent on their absolute values) such as classification, clustering, matching and change detection. Pansharpening products in remote sensing are mostly used for further processing in the above mentioned applications thus a new quality measure CMSC (translation invariant with respect to all parameters) proposed in [10] can be more suitable/justified

$$CMSC(x,y) = \left(1 - d_1\right) \cdot (1 - d_2) \cdot \rho,$$
$$d_1 = \frac{\left(\mu_x - \mu_y\right)^2}{R^2},\ d_2 = \frac{(\sigma_x - \sigma_y)^2}{\left(R/2\right)^2}, \tag{2}$$

where $R=2^8=255$ for 8bit data.

As the reference image is not available in pansharpening applications an ideal way to assess quality of pansharpening products would be to evaluate their impact in particular application by using reference/ground truth data of a given application. This way is very time and resource consuming thus not practical for the selection of a suitable method from maybe hundreds available. We have to note, that there exist some attempts to measure image quality without reference mostly based on gradients but they are not enough sensitive to fine differences which occur during pansharpening process. Thus the following two practical approaches have established over the past two decades. **The first approach** is based on the simulation (low pass filtering) of the input data with a further pansharpening of reduced resolution data (Wald's protocol) and then comparing fusion result with the original multispectral image using FR measures listed above. Unfortunately, there is no evidence that results/conclusions obtained in low resolution scale can be directly/automatically transferred to high resolution. **The second approach** is based on the comparison of fusion result with the two inputs of pansharpening: low resolution multispectral image and high resolution panchromatic image. For the comparison with original multispectral image the fusion result should be low pass filtered. Such assessment leads to two types of measures (set of measures) derived in low (spectral) and high (spatial) resolution scales. The same FR measures are used as mentioned above. We can mention some examples of such quality assessment sometimes called protocols e.g. Zhou's protocol (uses CC), QNR measure (uses UIQI), Khan's protocol (uses Q4 and UIQI), product of two measures (uses CC and UIQI) [7]. It is observed that it is quite difficult to rank methods using several measures thus sole or joint measures were proposed recently such as QNR [6], product of two measures [7] and JQM [8]. We propose to enhance a joint quality measure JQM [8] by **replacing SSIM with a newly introduced CMSC**. In this case the separate parts of JQM can be written in the following way. Quality for Low Resolution (QLR) is defined in reduced resolution space and compares multispectral images

$$QLR = \sum w_k \cdot CMSC(ms_k, (msf_k * lpf) \downarrow), \tag{3}$$

where (high resolution data)$\downarrow$ means subsampling of high resolution data to low resolution data. Quality for High Resolution (QHR) is defined in high resolution space

$$QHR = CMSC(pan, \sum w_k \cdot msf_k) . \quad (4)$$

Finally JQM is defined as a weighted sum of both measures

$$JQM = v_1 \cdot QLR + v_2 \cdot QHR, \; v_1 + v_2 = 1 . \quad (5)$$

4. EXPERIMENTS

We will illustrate our ideas concerning pansharpening quality assessment for optical remote sensing sensor IKONOS over Munich city in South Germany. In this section, we will compare different methods: GFF, HPFM, CS, GS, ATWT and Ehlers (see Sect. 2) using the proposed JQM and already known QNR joint quality measures. Values of both quality measures and corresponding separate measures are presented in Table 1 and Figs 1-4 for various interpolation and fusion methods and various parameter settings (cutoff frequencies of low pass filter).

First, only interpolated multispectral data are evaluated.

Table 1. Quality assessment of interpolation methods.

Method / Measure	NN	ZP	BIL	CUB
QLR	0.9999	0.9909	0.9815	0.9931
QHR	0.4279	0.5165	0.5192	0.5307
JQM	**0.7139**	**0.7537**	**0.7503**	**0.7619**
1-DL	0.8909	0.9105	0.8762	0.8959
1-DS	0.8573	0.8825	0.8806	0.8859
QNR	**0.7638**	**0.8035**	**0.7716**	**0.7939**

We see that all interpolation methods exhibit quite similar QLR values (Table 1). Similarly all methods have quite similar QHR values except NN. NN has very poor spatial quality what leads to low (poor) values of JQM for NN. Moderately oscillating values of separate measures QLR and QHR for other three methods result in slightly higher values of CUB. Analysis of QNR is more complex due to greater variability of its compound parts. 1-D_λ measure prefers ZP followed by CUB. BIL and NN seem to be the worst. 1-D_s measure behaves similarly to QHR. Thus, QNR value follows approximately results of separate measure 1-D_λ, finally underestimating BIL method (similarity of NN and BIL contradicts visual analysis). QNR founds NN as a worst method what corresponds quite well to JQM in this case. In total, it seems that both joint quality measures behave quite similarly except that QNR (1-D_λ) tends to underestimate BIL interpolation quality. Moreover, 1-D_λ measure appears to be more sensitive (exhibits higher variability).

QLR measure behaves as expected for GFF (methods 1-4) and HPFM (methods 5-8) in dependence of cutoff frequencies (Fig. 3) that is QLR increases with the increase of cutoff frequency (spectral quality). QHR selects methods 2 and 6 as the best, what corresponds quite well with visual analysis. Further JQM selects methods 3 and 7 with band dependent cutoff frequencies (Fig. 1), what is well supported by visual interpretation. Moreover, it seems that HPFM (faster variant of GFF) is better than GFF, maybe, due to the different interpolation method used. Thus both measures QHR and JQM are able to correctly select optimal cutoff frequencies for both methods.

Spectral measure 1-D_λ follows approximately the behavior of QLR for methods 1-8 (Fig. 4). Spatial measure 1-D_s follows the trend of 1-D_λ, what contradicts visual analysis. Such behavior of these two measures leads to the same trend of joint quality measure QNR in Fig. 2. Thus QNR is not able to select optimal cutoff frequencies for GFF and HPFM methods.

QLR of other methods (CS IHS (9), CS GS (10), ATWT (11) and Ehlers (12)) is lower than these of most filtering methods, whereas for QHR the opposite observation is valid. Finally, JQM of these methods (9-12) is lower than these of the best filtering methods (2-3, 6-7). QNR ranks methods 9-12 close to methods 1, 5 with high spatial quality. Only Ehlers (method 12) becomes high overall score.

In conclusion we mention one more observation or drawback of QNR limiting its practical usage. JQM values of any pansharpening method (Fig. 1) are higher than those of only interpolation methods (Table 1). In contrary, QNR values of all interpolation methods (Table 1) are higher than these of all pansharpening methods, except methods 4 and 8, whose quality as we know already is estimated wrongly.

5. CONCLUSIONS

The Joint Quality Measure JQM is proposed which is based on the new FR measure CMSC and performs comparison of a fusion result separately (QLR and QHR) with each of the inputs of pansharpening. It allows practical selection of optimal filtering parameters and comparison of different pansharpening methods. The results are well supported by visual analysis and existing experience.

Already known QNR measure tends to underestimate the quality of BIL interpolation. Additionally, its spatial part 1-DS seems to be not able to correctly rank filtering based fusion methods in dependence of the filtering parameter (quality for large parameter values is overestimated). Moreover, 1-DS overestimates quality of all interpolation methods when compared with almost all fusion methods. Exceptions are filtering based methods with large parameters values whose quality is again overestimated as already stated above. The cause of these drawbacks of 1-DS can be its wrong/incorrect usage/definition (bands with different spectral ranges are compared in this measure).

6. REFERENCES

[1] J. Yang, and J. Zhang, "Pansharpening: from a generalized model perspective," *International Journal of Image and Data Fusion* 5(4), pp. 285-299, 2014.

[2] Q. Xu, Y. Zhang, and B. Li, “Recent advances in pansharpening and key problems in applications,” *International Journal of Image and Data Fusion* 5(3), pp. 175-195, 2014.
[3] B. Aiazzi, L. Alparone, S. Baronti, and A. Garzelli, “Quality assessment of pansharpening methods and products,” *IEEE Geoscience and Remote Sensing Society Newsletter*, pp. 10-18, December 2011.
[4] A. Makarau, G. Palubinskas, and P. Reinartz, “Analysis and selection of pan-sharpening assessment measures,” *Journal of Applied Remote Sensing* 6(1), 063548, pp. 1-20, 2012.
[5] A. Makarau, G. Palubinskas, and P. Reinartz, “Selection of numerical measures for pan-sharpering assessment,” Proc. of IGARSS, Munich, Germany, IEEE, pp. 2264-2267, 2012.
[6] L. Alparone, B. Aiazzi, S. Baronti, A. Garzelli, F. Nencini, and M. Selva, “Multispectral and Panchromatic Data Fusion Assessment without Reference,” *Photogrammetric Engineering & Remote Sensing* 74(2), pp. 193-200, 2008.
[7] C. Padwick, M. Deskevich, F. Pacifici, and S. Smallwood, “WorldView-2 Pan-Sharpening,” Proc. of ASPRS, April 26-30 2010, San Diego, California, USA, pp. 1-13, 2010.
[8] G. Palubinskas, “Quality assessment of pan-sharpening methods,” Proc. of IGARSS, Québec city, Canada, IEEE, pp. 2526-2529, 2014.
[9] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, “Image Quality Assessment: From Error Visibility to Structural Similarity,” *IEEE Trans. Image Processing* 13(4), pp. 600-612, 2004.
[10] G. Palubinskas, “Mystery behind similarity measures MSE and SSIM,” Proc. International Conference on Image Processing, 27-30 October, 2014, Paris, France, IEEE, pp. 575-579, 2014.
[11] Z. Wang, D. Ziou, C. Armenakis, D. Li, and Q. Li, “A comparative analysis of image fusion methods,” *IEEE Trans. on Geoscience and Remote Sensing* 43(6), pp. 1391-1401, 2005.
[12] B. Aiazzi, S. Baronti, F. Lotti, and M. Selva, “A comparison between global and context-adaptive pansharpening of multispectral images,” *IEEE Geoscience and Remote Sensing Letters* 6(2), pp. 302-306, 2009.
[13] G. Palubinskas, and P. Reinartz, “Multi-resolution, multi-sensor image fusion: general fusion framework,” Proc. of JURSE, Munich, Germany, IEEE, pp. 313-316, 2011.
[14] G. Palubinskas, “Fast, simple and good pan-sharpening method,” *Journal of Applied Remote Sensing* 7(1), 073526, pp. 1-12, 2013.
[15] S. Klonus, and M. Ehlers, “Image fusion using the Ehlers spectral characteristics preservation algorithm,” *GIScience & Remote Sensing* 44(2), pp. 93–116, 2007.
[16] B. Aiazzi, L. Alparone, S. Baronti, and A. Garzelli, “Context-Driven Fusion of High Spatial and Spectral Resolution Images Based on Oversampled Multiresolution Analysis,” *IEEE Trans. on Geoscience and Remote Sensing* 40(10), pp. 2300-2312, 2002.
[17] M. J. Canty, *Image analysis, classification and change detection in remote sensing: with algorithms for ENVI/IDL and Python*, Third revised edition, Taylor & Francis, CRC Press, 2014.

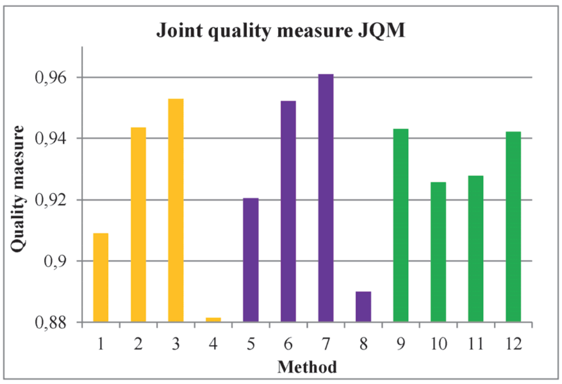

Figure 1. JQM for 6 methods and their different parameter settings[1].

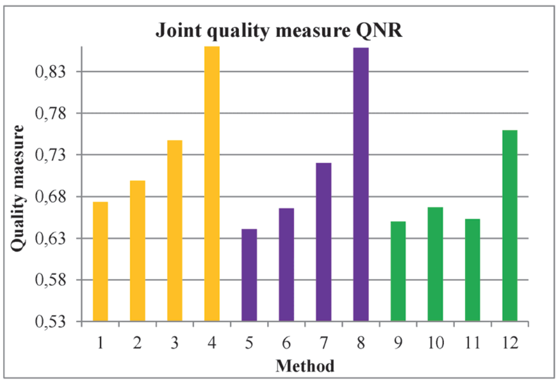

Figure 2. QNR for 6 methods and their different parameter settings.

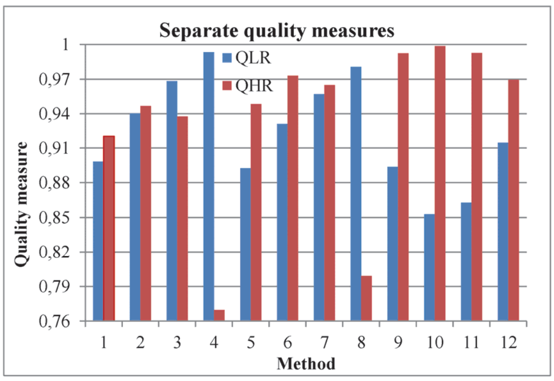

Figure 3. Separate measures for JQM (Fig. 1).

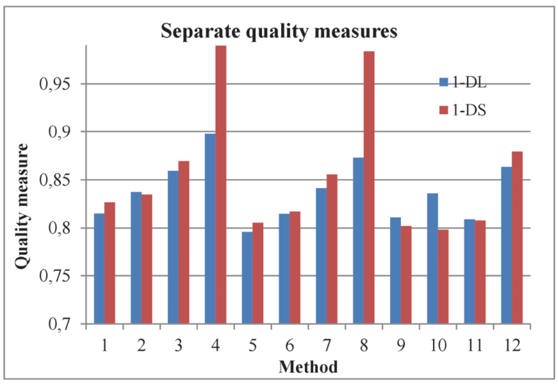

Figure 4. Separate measures for QNR (Fig. 2).

[1]List of methods: 1-GFF (cutoff frequency 0.05), 2-GFF (0.15), 3-GFF (0.3,0.2,0.1,0.1), 4-GFF (0.7), 5-HPFM (0.05,BIL), 6-HPFM (0.2,BIL), 7-HPFM (0.3,0.25,0.15,0.1,BIL), 8-HPFM (0.7,BIL), 9-CS IHS (BIL), 10-CS GS, 11-ATWT, 12-Ehlers (0.15,0.15,CUB).