SELECTION OF NUMERICAL MEASURES FOR PAN-SHARPENING ASSESSMENT

Aliaksei Makarau, Gintautas Palubinskas, and Peter Reinartz

Photogrammetry and Image Analysis, Remote Sensing Technology Institute
German Aerospace Center DLR Oberpfaffenhofen,
D-82234 Wessling, Germany

ABSTRACT

Different tasks of multispectral image analysis and processing require specific properties of input pan-sharpened multispectral data such as spectral and spatial consistency. Generally, the quantitative measures for pan-sharpening assessment were taken from other topics of image processing (e.g. image similarity indexes). All these measures are widely employed for this task but the applicability basis of these measures is not checked and proven. In this paper a comparison of pan-sharpening assessment measures for remote sensing is carried out on specially generated pan-sharpened images. Performed statistical analysis on the assessment measures allows to select the measures which are most sensitive to the pan-sharpened imagery quality and these measures are recommended for use.

Index Terms— Pan-sharpening, quality assessment

1. INTRODUCTION

Pan-sharpened remote sensing data have many areas of application, therefore different requirements are posed on the pan-sharpened data quality. The requirements can be on spectral consistency, spatial consistency or on both together. Spectral consistency assumes that the pan-sharpened image has increased spatial resolution with spectral properties of the original image. Spatial consistency assumes that "A high spatial quality merged image is that which incorporates the spatial detail features present in the panchromatic image and missing in the initial multispectral one" [1]. The ideal case is the highest spectral and spatial consistency together.

In many cases several measures can give contradictory results and the decision on the pan-sharpening quality is difficult to make. Such contradictory results may be caused by the fact that the measures are inappropriate for such use. Therefore the question on the applicability of the measures should be made clear.

In this paper a comparison of pan-sharpening assessment measures for remote sensing is carried out on a specially generated test set of images. The test set is composed of pan-sharpened images, produced with different quality (spectral and spatial consistency). Analysis of variance and pairwise comparison statistical methods are performed on the assessment measures. The pan-sharpening assessment measure is required to be sensitive to the pan-sharpened imagery quality change (i.e. able to separate imagery with different quality) as well as sensitive to the increase or decrease of the image quality (i.e. provides the increase or decrease of the assessment score). The measures that are most sensitive to the quality change (according to statistical assessment results) are recommended for use.

[1]We would like to thank European Space Imaging (EUSI) for the collection and provision of DigitalGlobe WorldView-2 and IKONOS-2 data.

2. ASSESSMENT AND SELECTION OF MEASURES

A pan-sharpening method may provide a perfect spectral consistency together with a poor spatial consistency and vice versa. Therefore, to make a proper assessment of a fusion result, assessment of both spectral and spatial consistency should be performed. An assessment measure (spectral or spatial consistency) should calculate a score according to the pan-sharpened image quality. The assessment measure should be sensitive to change (monotonous increase or decrease) of pan-sharpened image quality. The higher the quality, the higher the calculated score of the measure and vice versa. A numerical measure can be assessed using test data, i.e. a test set of images. Variation of image quality in the test set allows to analyze the sensitivity of the measure using statistical methods.

The most known and popular similarity measures used for spectral consistency assessment are: Spectral Angle Mapper, SAM (calculated as the angle between two vectors, which are composed using the pixel values of the compared multispectral images); Structural SIMilarity SSIM [2] or extended SSIM - Q4, (correlation, contrast, and luminance similarity between two images are used to calculate one similarity value); ERGAS [3] (similarity measure for multispectral images, based on the mean squared error estimator); Zero mean normalized cross-correlation, usually named as CORR.

Up to now not many papers deal with spatial consistency assessment. Almost all the methods use a single scale edge detector (Gradient, Laplacian, Sobel edge detector) and an evaluation metric to calculate the distance between the

edge maps (usually correlation coefficient) [4]. For example, the High Pass Correlation Coefficient (HPCC) employs Laplacian and normalized correlation. Here the comparison is made between the fused bands and the corresponding panchromatic image. Another approach calculates the percentage of true and false edges introduced into the fused band using the Sobel edge detector [4]. Several works on fusion use the SSIM and ERGAS measures for spatial consistency assessment [1, 5] (panchromatic image is used as the reference instead of a spectral band, the measures are labeled as as SSIM_PAN and ERGAS_PAN).

An additional measure for spatial consistency assessment was proposed for use in [6]. This measure uses phase congruency (PC) [7] for feature extraction from pan-sharpened image. Invariance to intensity and contrast change as well as multiscale nature of this measure allows to obtain more confident assessment comparing to single-scale edge detectors.

2.1. Test data generation

Medium and high resolution (Landsat 7 ETM+, IKONOS, and WorldView-2) spaceborne imagery was used for generation of the test sets (one test set is produced for each sensor). The images were obtained in different parts of the Earth and have different land cover classes, such as urban, rural, agricultural areas, forest and water regions to represent a high variety. Two scenes for each satellite were chosen. Landsat 7 ETM+ images (8-bit) were acquired at 7-th July 1999, and at 13-th September 1999 for the areas of San Jose city (USA) and Plattling town (Germany). IKONOS images (11-bit) were acquired at 15-th July 2005, 10:28 GMT, and at 24-th July 2004, 09:25 GMT for the areas of Munich city (Germany) and Athens city (Greece). WorldView-2 (16-bit) images were acquired at 12-th July 2010, 10:30 GMT, and at 10-th December 2009, 10:30 GMT for the areas of north of Munich city (Germany) and Rome city (Italy). Ten non-overlapping tiles were taken from the acquired images (multispectral and panchromatic) for each sensor. The size is 2000×2000 for panchromatic and 500×500 for multispectral IKONOS, WorldView-2, or 1000×1000 for multispectral Landsat 7 ETM+.

The GFF pansharpening method [8] is used for the test set generation. The GFF method similarly as the General Image Fusion (GIF) method [9] shows that many pan-sharpening methods are quite similar and can be described as special cases of more general fusion methods. This method allows to control the quality of produced pan-sharpened image by varying the parameter set. The GFF has a parameter hf, which varies in the range [0, 1] and controls the proportionality (0%-100%) of high-frequency panchromatic image data to be added to low-resolution spectral image. The high-frequency information is extracted using the Butterworth filter. Variation of this parameter allows to create fused images with desired quality: the higher the hf value, the more high-frequency data is added, and the higher spatial (lower spectral) consistency, and vice versa.

The nonoverlapping tiles are pan-sharpened by the GFF method with five values for the parameter hf (hf=0.95, 0.90, 0.85, 0.80, and 0.75, i.e. 95%, 90%, 85%, 80%, and 75% of high frequency panchromatic image data is added). Five groups of fused images (each group consists of ten pan-sharpened tiles) are generated for each test set (Figure 1, Step 1). To show that the GFF performs pan-sharpening with a competitive quality, the same image tiles are pan-sharpened by the ARSIS fusion [10] (ARSIS is used for comparison).

Figure 2 illustrates an example of WorldView-2 image pan-sharpening by the GFF method (a single band, green, 510-580 nm is presented for comparison). The GFF was run five times with different parameter value (hf=0.95, 0.90, 0.85, 0.80, and 0.75). The spatial consistency of the fused image decreases from the 95% of added high frequency data (Figure 2(a)) to 75% (Figure 2(e))), while the spectral consistency increases. Figure 2(f) contains the ARSIS fusion. The assessment scores (ERGAS, CORR, and HPCC) illustrate that the ARSIS fusion is comparable to the GFF fusion with 90% and 85% of added high frequency data (see Table 1).

Table 1. ERGAS, CORR, and HPCC spectral consistency measures calculated for the images shown in Figure 2 (all bands employed)

Figure	pan-sharpening method	ERGAS	CORR	HPCC
2(a)	GFF 95%	2.2535	0.9374	0.9530
2(b)	GFF 90%	2.0774	0.9493	0.9670
2(c)	GFF 85%	1.9103	0.9597	0.9567
2(d)	GFF 80%	1.7976	0.9664	0.9326
2(e)	GFF 75%	1.7309	0.9702	0.9043
2(f)	ARSIS	2.1083	0.9585	0.9900

2.2. Statistical assessment

Measures on spectral consistency are performed using Wald's protocol. To obtain one numeric score for a multispectral image the mean value is taken on the scores calculated for the channels. Five groups of assessment scores (also named as: 95%, 90%, 85%, 80%, and 75%) (Figure 1, Step 2) are calculated and used to assess the sensitivity of the measure to the quality change (the first test) and to estimate and analyze the trend of the measure (increase or decrease, e.g. the measure score change from the 95% to 90%, from the 90% to 85%, and so forth; the second test).

Sensitivity of a measure to the quality variation is assessed using the Kruskal-Wallis one-way analysis of variance [11] on the scores calculated for the 95%, 90%, 85%, 80%, and 75% pan-sharpened images (Figure 1, Step 3). The Kruskal-Wallis test is a non-parametric alternative to the one-way Analysis of variance (ANOVA) and used for testing equality of population medians among groups. Sensitivity of a measure to the monotonous change of the quality is assessed using pairwise comparisons (pairwise one tailed Wilcoxon rank sum test). The test is performed on each of the following four pairs of

the score groups: 1) 0.95% and 0.90%; 2) 0.90% and 0.85%; 3) 0.85% and 0.80%; 4) 0.80% and 0.75% (Figure 1, Step 3). In the case of a quality increase it is expected that the median of the 0.95% group is less than the median of the 0.90% group, and so forth for the measure with minimal score equal to zero and the ideal value equal to some value.

3. RESULTS AND DISCUSSION

The spectral consistency of the produced groups of pan-sharpened images (95%, 90%, 85%, 80%, and 75%) is monotonously increasing, while the spatial consistency is monotonously decreasing (less high frequency data is added). Therefore the numerical scores of the measures SSIM, CORR (spectral consistency), and ERGAS_PAN (spatial consistency) are expected to increase, while the ERGAS and SAM scores are expected to decrease. The SSIM_PAN, CORR_PAN, HPCC, and PC (spatial consistency) scores are also expected to decrease (less high frequency data is added).

For Landsat 7 ETM+ pan-sharpening assessment the ERGAS, CORR (spectral consistency), and the PC, SSIM_PAN (spatial) measures are preferable according to separability of the pan-sharpened imagery and estimated trends. For IKONOS the CORR (spectral), and PC (spatial) measures are preferable. The measures show regular estimation of the quality trend (CORR score increase) and PC score decrease. For WorldView-2 the CORR (spectral), and SSIM_PAN, and PC (spatial) measures are preferable. The PC was run with default values of the parameter set. Nevertheless, this measure provides required characteristics on the test data with different spatial resolution. The boxplots illustrating a wrong trend (e.g. 3(a), and 3(h)) show that the measures do not estimate the quality change properly.

4. CONCLUSIONS

Statistical analysis revealed that not all the widely employed measures calculate accurate regular results of pan-sharpened imagery assessment and distortion of calculated scores can appear. Spectral consistency assessment should employ ERGAS and Normalized Correlation measures (ERGAS and Normalized Correlation for Landsat 7 ETM+, Normalized Correlation for IKONOS and WorldView-2). Spatial consistency assessment should employ Phase Congruency and SSIM measures (Phase Congruency and SSIM for Landsat 7 ETM+ and WorldView-2, Phase Congruency for IKONOS). SAM provides stable assessment scores of the imagery irrespectively to varying spectral and spatial consistency, therefore it should be used with caution. Phase Congruency shows a good separability of pan-sharpened imagery in the sense of spatial consistency, and sensitivity to the trend of the quality change (for all used sensors data).

5. REFERENCES

[1] M. González-Audícana, X. Otazu, O. Fors, and A. Seco, "Comparison between Mallat's and the á trous discrete wavelet transform based algorithms for the fusion of multispectral and panchromatic images," *Int. Journal of Remote Sensing*, vol. 26, no. 3, pp. 595–614, 2005.

[2] Zhou Wang, Alan C. Bovik, and Hamid R. Sheikh, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[3] L. Wald, T. Ranchin, and M. Mangolini, "Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images," *PE & RS*, vol. 63, no. 6, pp. 691–699, 1997.

[4] P. Pradhan, R. King, N. Younan, and D. Holcomb, "Estimation of the number of decomposition levels for a wavelet-based multiresolution multisensor image fusion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 12, pp. 3674–3686, 2006.

[5] M. Lillo-Saavedra, C. Gonzalo, C. Arquero, and E. Martinez, "Fusion of multispectral and panchromatic satellite sensor imagery based on tailored filtering in the Fourier domain," *International Journal of Remote Sensing*, vol. 26, no. 6, pp. 1263–1268, 2005.

[6] A. Makarau, G. Palubinskas, and P. Reinartz, "Analysis and selection of pan-sharpening assessment measures," *Journal of Applied Remote Sensing (accepted)*, 2012.

[7] P. Kovesi, "Image features from phase congruency," *Videre: A Journal of Computer Vision Research*, vol. 1, no. 3, pp. 2–26, 1999.

[8] G. Palubinskas and P. Reinartz, "Multi-resolution, multi-sensor image fusion: general fusion framework," in *Joint Urban Remote Sensing Event (JURSE), 2011*, april 2011, pp. 313–316.

[9] 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. 1391–1402, 2005.

[10] T. Ranchin and L. Wald, "Fusion of high spatial and spectral resolution images: The ARSIS concept and its implementation," *Photogrammetric Engineering & Remote Sensing*, vol. 66, no. 1, pp. 49–61, 2000.

[11] David J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures, Second Edition*, Chapman & Hall/CRC, 2000.

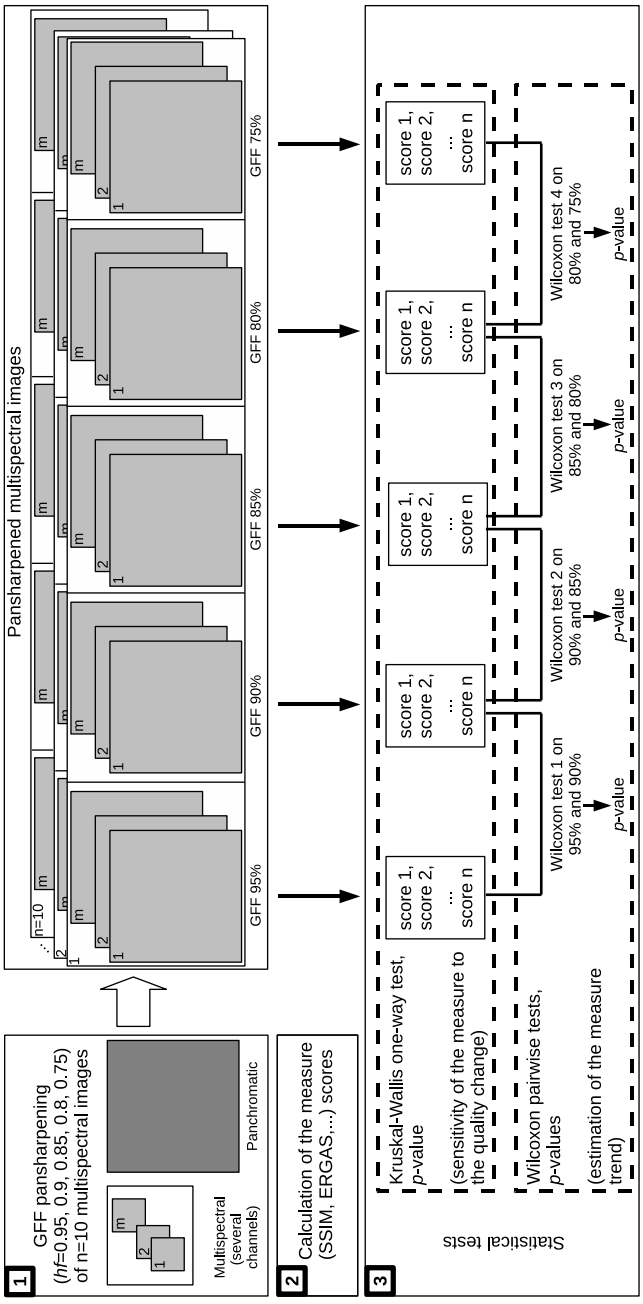

Fig. 1. Statistical assessment of pan-sharpening assessment measures. First, the multispectral images ($n = 10$ multispectral images with m bands) are pan-sharpened by the GFF method five times with different parameter hf. Second, the numerical scores are calculated by the assessment measures (five groups of numerical scores, each group consists of $n = 10$ scores) are produced. Third, the statistical tests are performed: Kruskal-Wallis one-way test on the five groups of numerical scores, pairwise Wilcoxon test is performed on the pairs of the groups

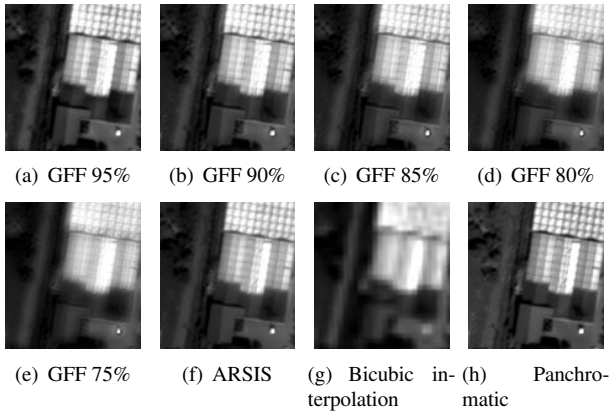

Fig. 2. GFF pan-sharpening (run with varying hf=0.95, 0.90, 0.85, 0.80, and 0.75) of WorldView-2 image is shown (green band is used for easier visual comparison). The quality of the ARSIS fusion (f) is comparable with the GFF 90% (b)

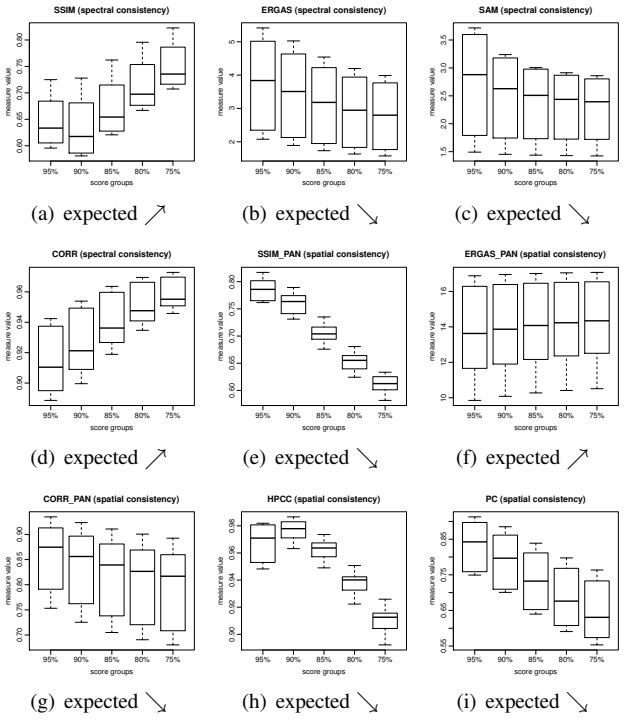

Fig. 3. Boxplots of the measures scores run on WorldView-2 test data (95%, 90%, 85%, 80%, and 75%): (a) SSIM, (b) ERGAS, (c) SAM, (d) CORR, (e) SSIM_PAN, (f) ERGAS_PAN, (g) CORR_PAN, (h) HPCC, (i) PC.