# A REVIEW ON SPATIAL QUALITY ASSESSMENT METHODS FOR EVALUATION OF PAN-SHARPENED SATELLITE IMAGERY

F. DadrasJavan<sup>1,\*</sup>, F. Samadzadegan<sup>1</sup>, S. Mehravar<sup>1</sup>, A. Toosi<sup>1</sup>

<sup>1</sup> Department of Geomatics, University College of Engineering, University of Tehran, Tehran, Iran-(fdadrasjavan, samadz, sorooshmehravar, ahmadtoosi71)@ut.ac.ir

## Commission VI, WG VI/4

KEY WORDS: Spatial Quality, Assessment Methods, High-Resolution Satellite Imagery, Image Fusion, Pan-sharpening

## **ABSTRACT:**

Nowadays, high-resolution fused satellite imagery is widely used in multiple remote sensing applications. Although the spectral quality of pan-sharpened images plays an important role in many applications, spatial quality becomes more important in numerous cases. The high spatial quality of the fused image is essential for extraction, identification and reconstruction of significant image objects, and will result in producing high-quality large scale maps especially in the urban areas. This paper introduces the most sensitive and effective methods in detecting the spatial distortion of fused images by implementing a number of spatial quality assessment indices that are utilized in the field of remote sensing and image processing. In this regard, in order to recognize the ability of quality assessment indices for detecting the spatial distortion quantity of fused images, input images of the fusion process are affected by some intentional spatial distortions based on non-registration error. The capabilities of the investigated metrics are evaluated on four different fused images derived from Ikonos and WorldView-2 initial images. Achieved results obviously explicate that two methods namely Edge Variance Distortion and the spatial component of QNR metric called Ds are more sensitive and responsive to the imported errors.

# 1. INTRODUCTION

Most of the modern topographic earth observation satellites, such as WorldView-2, WorldView-3, QuickBird, Ikonos, Pleiades and Kompsat, provide a panchromatic (PAN) image at a high spatial resolution and a multi-spectral (MS) image at a high spectral resolution, but these images contain lower spectral and spatial information, respectively (Snehmani et al., 2017). Due to technical limitations, sensors cannot produce unique images that have high spatial and high spectral resolution simultaneously (Nikolakopoulos and Oikonomidis, 2015; Vivone et al., 2015). Considering that many applications of satellite imagery require both high spatial and spectral quality, the process of sharpening MS images by PAN images called pan-sharpening has been raised (Wald et al., 1997; Vijayaraj et al., 2004). Thus far, several methods and algorithms for pansharpening high-resolution satellite imagery are presented in the literature (Thomas and Wald, 2004; Alparone et al., 2007; Amolins et al., 2007; Ehlers et al., 2010). In the procedure of pan-sharpening methods, the input images will change and ultimately new data with different characteristics will be obtained. Because of the different fusion methods and the different qualities they provide, evaluation of the quality of images fused by different fusion methods is so important. On the other hand, among all proposed quality assessment (QA) methods, only a few of them consider the spatial features of the image. Spectral quality has a higher priority in some applications, such as interpretation and classification, while in applications such as extraction, identification, and reconstruction of image objects, spatial quality plays a more important role (Javan et al., 2013).

#### 2. PROPOSED METHOD

#### 2.1 Process Description

One of the spatial distortions associated with the pansharpening process is the non-registration problem of the input images. In this study, miss-registration error is used to generate spatially distorted fused images. The similarity of the missregistration spatial distortion to spatial distortions that occur during the pan-sharpening process is presented by showing multiple crops of these distorted images in Figure 3 and 4. Therefore, this error can be a good representation of the spatial distortions happen during the fusion process. For the missregistration error, the corresponding pixels of the MS image and PAN image are displaced relative to each other. Misplacements or shifts of the input MS image compared to the PAN image, are selected equal to 0.5 pixels, 1 pixel, and 1.5 pixels, and are implemented in both horizontal and vertical directions. In this way, there will be four sets of images for each selected data, including an error-free set and three image sets that have a deliberate error. In the final step, the stored data sets have been

By increasing the resolution of satellite images in recent years, it is possible to produce large scale maps of the fused images, so the spatial QA of the pan-sharpened images has undeniable importance. To overcome the lack of robust spatial quality metrics in the field of image fusion, this paper has collected and reviewed some well-known indices proposed in the image processing domain. Ultimately, all investigated QA methods are compared and new optimal metrics that are more capable of detecting spatial distortions are identified.

<sup>\*</sup> Corresponding author.

used to determine the ability and sensitivity of spatial QA methods and their ranking.

## 2.2 Fusion Methods and Quality Assessment Metrics

In this paper, 15 indices namely Correlation Coefficient (CC), Edge extraction CC (Edge CC), Spatial CC (SCC), Mean Gradient (MG), Ds component of QNR index (QNR-Ds), Spatial Frequency (SF), Entropy, Standard Deviation (STD), Accuracy component of Edge Preservation Ratio (EPRa), Robustness component of Edge Preservation Ratio (EPRr), Edge variance distortion (EVD), Gradient Magnitude Similarity (G-M-SIM), Gradient Direction Similarity (G-D-SIM), Edge similarity for screen content image (ESIM), and Just Noticeable Blur (JNB) are used for the assessment step. The requirement of mentioned metrics to reference images (RI) and their references are presented in Table 1.

Metric	RI	References
	Requirement	
CC	Yes	(Wang et al., 2004)
Edge CC	Yes	(Ehlers et al,. 2010)
SCC	Yes	(Zhou et al,. 1998; Khan et al., 2008)
MG	No	(Ryan et al, 2003)
QNR-Ds	Yes	(Pálsson, 2013; Mandhare et al., 2013; Palubinskas, 2015)
SF	No	(Jagalingam and Hegde, 2015)
Entropy	No	(Naidu, 2010; Alimuddin et al., 2012)
STD	No	(Al-Wassai and Kalyankar., 2012)
EPRa	Yes	(Yu et al., 2014)
EPRr	Yes	(Yu et al., 2014)
EVD	Yes	(Qi et al., 2014)
G-M-SIM	Yes	(Xue et al., 2013)
G-D-SIM	Yes	(Ni et al., 2017)
ESIM	Yes	(Ni et al., 2017)
JNB	No	(Ferzli and Karam., 2009)

Table 1. Investigated spatial quality metrics.

In the following, a concise description of each QA metric is presented.

The CC index is a popular criterion for estimating the amount of similarity between the two images. Although CC is a distinct index and can be implemented on the greyscale values of images, it has been utilized in many other metrics for comparing the extracted information of images. One example of this kind is SCC metric in which the high-frequency details of PAN and fused images are firstly extracted by a high-pass 2-D filter that is the Laplacian filter commonly; then, correlation calculation is used for similarity determination of these detail maps. In the case of Edge CC, the similarity value will be computed on the edge maps of PAN and fused images using the correlation definition.

The mean value of gradient magnitude in the MG metric is a known way for estimating the amount of spatial information in an image in which the higher values indicate the higher spatial information and vice versa. In this paper, the Sobel operator is applied among all other gradient operators because of its efficiency and simplicity of computation.

The spatial component of the QNR index  $(D_S)$  is computed through similarity measurements of couples of scalar images executed by employing the universal image quality index (UIQI) metric. Moreover, the algorithm of QNR uses both PAN and MS images in the computations and can model the quality by combining three factors, including loss of correlation, luminance distortion and contrast distortion (Pálsson, 2013). According to the Equation 1 which represents the  $D_S$  component of QNR metric, Q is the UIQI index,  $MS_K$  is the low-resolution Kth MS band,  $PAN_L$  indicates the PAN image which is degraded to the resolution of MS image. Also,  $Fused_K$  is the Kth band of pan-sharpened image and is defined to be compared with the original PAN image and N denotes the number of MS bands.

$$D_{S} = \frac{1}{N} \sum_{K=1}^{N} |Q(MS_{K}, PAN_{L}) - Q(Fused_{K}, PAN)|$$
(1)

The SF index does not require a reference image and calculates the level of spatial information in both directions of rows and columns by comparing the adjacent greyscale values of the fused image.

The entropy of a pan-sharpened image demonstrates the randomness level of image content and is sensitive to the noises.

SD index can measure the contrast in fused images and higher SD values are commonly anticipated in high-contrast images (Naidu, 2010).

Inspired by the structural similarity equation (Wang et al., 2004), the G-M-SIM and G-D-SIM metrics can be calculated after the computation of gradient magnitudes and gradient directions of PAN and fused image (Xue et al., 2013). Considering the achieved gradient magnitudes of PAN and fused images namely  $m_{PAN}$  and  $m_{Fused}$ , calculation of gradient magnitude similarity will be done at the location *i* using Equation 2, where c is a positive constant that supplies numerical stability.

$$G - M - SIM(i) = \frac{2m_{PAN}(i)m_{Fused}(i) + c}{m_{PAN}^2(i) + m_{Fused}^2(i) + c}$$
(2)

The JNB metric can detect the blurriness level of images with different contents and is designated in such a way to model the human visual system and not require any reference images. Based on this method, the blur amount is dependent on local contrast and edge width values (Ferzli and Karam., 2009).

ESIM is an edge-based QA method which is composed of three image features, including edge contrast, edge direction, and edge width. After extracting the features from the PAN and fused images, the overall score of the metric is computed by pooling the calculated similarity values of the mentioned features.

The other investigated spatial QA indices of this paper are based on edge analysis. The general EPR method consists of accuracy (EPRa) and robustness (EPRr) components which the former refers to the ratio of preserved edge pixels in the distorted image to the number of edge pixels in the reference image and the latter indicates the ratio between the number of preserved edge pixels to the number of edge pixels in the distorted image. The EVD metric takes into account the edge differences and compares the edge map of the pan-sharpened image with the edge map of the PAN image to count the number of eliminated and extra edge pixels in the fused image. It also allocates a higher weight factor to the areas containing more spatial details by dividing the reference image into the flat blocks and high detail blocks.

For the fusion step, the un-decimated Weighted Wavelet Intensity (WWI) (Zhang & Hong, 2005) method, which is one of the most acceptable methods in the spatial quality component, has been used (Javan et al., 2018). In addition, Matting Model pan-sharpening (MMP) (Kang and Benediktsson, 2013) and Band Dependent Spatial Detail (BDSD) (Garzelli et al., 2008) fusion methods that provide lower spatial quality than WWI method are selected. In fact, the purpose is to prove that the spatial distortion occurred during the fusion process is similar to the distortion caused by the non-registration error and thereby, the evaluation results can be generalized to all applications of fused images. Diagram of the proposed scheme including the general steps used in this paper is shown in Figure 1.



Figure 1. Diagram of the proposed strategy for determination of the optimal spatial indices in image fusion quality assessment.

## 3. EXPERIMENTS AND RESULTS

# 3.1 Data Set

In this paper, four data sets provided by WorldView-2 (WV2) and Ikonos sensors have been used and they are selected in such a way to include all land covers such as buildings, road, and vegetation. Therefore, two of the case studies, namely Washington DC-1 (Figure2) and Melbourne (Figure3) are selected from urban areas and contain high amount of details while the other two case studies namely Washington DC-2 (Figure4) and Sichuan (Figure5) are widely covered by vegetation and flat areas and can be considered as non-urban images. The data sets 1 and 3 are the two parts of a unique greater image acquired from the Washington city, but they are thoroughly different in the type of image contexts they cover.

This variety of image context will provide a more reliable comparison of QA metrics in the result section. The size of all image sets is great enough to cover an extensive region including man-made and natural structures.



Figure 2. Data set 1, Washington DC-1 image set.



Figure 3. Data set 2, Melbourne image set.



Figure 4. Data set 3, Washington DC-2 image set.



Figure 5. Data set 2, Sichuan image set.

Images provided by WV2 sensor have eight spectral bands and a PAN band, which have a resolution of 2 and 0.5 meters, respectively. The Ikonos sensor provides an MS image consisting of 4 spectral bands and a PAN band which their resolution is 1 and 4 meters, respectively.

#### 3.2 Pan-sharpened images

Some image crops of the error-free fused images and deliberate distorted pan-sharpened images are illustrated in Figure 6 and 7. The similarity of the distortions occurs in MMP and BDSD fusion methods to distortions caused by the non-registration error (Figure 6 and 7) is shown in Figure 8. As shown in Figure 6 and 7, the more non-registration error of the input images rises, the more distortion appears in the fused images. The image crops are selected from the areas containing spatial details, so any changes in image content in the presence of deliberate error will be clarified this way.

According to Figure 6 and 7, the blurring effects, edge displacements, spectral and spatial distortions could be seen that can simulate the ordinary fusion distortions.



Figure 6. Image crops of error-free and distorted fused images, pan-sharpened by WWI method.



Figure 7. Image crops of error-free and distorted fused images, pan-sharpened by WWI method.

As can be seen, the generated distortions have occurred on the spatial details of the images and their surrounded pixels. On the one hand, some common spatial distortions that happen in the process of low spatial quality fusion algorithms are illustrated in Figure 8 to approve the generalization potential of the proposed scheme of this paper.



Figure 8. Distortions caused by low spatial quality fusion methods.

## 3.3 Comparative Charts for Quality Assessment Methods

By analyzing the responses of the QA indices to the deliberate spatial distortions of fused images, their sensitivity and ability to identify the spatial distortion are ranked and the bests are presented. In this regard, all metric results are normalized so they could be comparable to each other. Comparative charts of metric results are illustrated in Figures 9, 10, 11 and 12. According to the figures, the horizontal axis represents the amount of registration error which is imported into the input MS image of the fusion process. In other words, the original image is referred to the properly fused image using the registered images. In other words, the original image refers to the properly fused image with the registered image sets. The other three categories namely 0.5 pixels, 1 pixel, and 1.5-pixel registration error are indicating the pan-sharpened images produced by the corresponding non-registered input data sets. The names of metrics and their corresponding colors are presented on the right side of the figures while the left side is a range of decimal numbers between zero and one and is a representation to estimate the ability of QA indices in response to the spatial distortions.

For the normalization step, the result value of the evaluation calculated for the proper pan-sharpened image is considered as the maximum value and thereby, other assessment results of the distorted fused images are normalized based on the mentioned maximum quantity. This action will help to determine the efficiency and ability of each metric by measuring the declining amount of QA results in the case of distorted images.





Figure 10. Comparative charts of metrics on data 2.







The degradation rate of the various metrics in response to intentional distortion is shown in Figures 9, 10, 11 and 12 and descending graphs show more robust indices.

#### 4. DISCUSSION AND CONCLUSIONS

According to the comparisons, it can be concluded that the QNR-Ds evaluation index has shown the best performance among all implemented metrics. In addition, the Edge Variance Distortion index has been able to estimate spatial distortion very well. As it was alluded before, it is anticipated that the more efficient and sensitive metrics provide more descending graphs in the presence of errors and vice versa. Regarding the derived charts for the four data sets, most of the QA indices indicate a slight decline of results while the imported non-registration errors are increased and this issue represents the less sensitivity of their corresponding metrics. The reason for the different performance of QA methods could be found in several main issues. Firstly, it is important to investigate the response of the indices that are dependent on reference images and compare them with no-reference methods which can lead to clarifying the range of sensitivity of each group. On the other side, since every single method utilizes one or several unique spatial features to obtain the spatial quality estimation of pan-sharpened images, it is expected to yield a different level of sensitivity for each metric. The diverse used algorithms and features in the metrics directly influence the acquired results. It is noticeable that the QNR metric is the only method among the indices which takes into account the effect of MS image in spatial evaluation. A reason for the better performance of the QNR index can be the spectral consideration which is not implemented in the algorithm of other spatial indices. There is some information in the pan-sharpened image that is created during the fusion process because of the color insertion and the contrast they produce. In general, this type of information is related to the existence of a considerable amount of spatial details in the MS image. Hence, incorporating the spectral features into the algorithm of spatial metrics will provide a better result for spatial evaluations. It is also obvious from the results that QA methods which require reference images have shown a better response to spatial distortions. Contrarily, negligible changes in the results of methods that work without reference images indicate their low sensitivity to spatial distortion. There is not any no-reference metric in the top five methods and a significant gap is seen between their graphs and other graphs. While the lack of ideal reference images has been a challenging topic for evaluations, we recommend using QA methods that require the PAN image as the alternative reference image instead of no-reference metrics. Even though there are some well-known no-reference metrics in the area of QA, they mostly can not thoroughly distinguish the noises and distortions from spatial details of an image that results in their less sensitivity to the fusion distortions.

All depicted responses of mentioned metrics have almost the same trend in all data sets which proves the reliability of results since they do not depend on the type of data sets. However, the results of data set 1 and 3 are more similar to each other among all case studies, because they have been taken from a single and great satellite image set.

Another key point of the achieved results is the superiority of QA methods based on edge analyses because they have produced more descending results in comparison with other metrics. While there are two non-urban data sets in the implementations, this trend is still the same in the shown result graphs of metrics as four out of the top five best metrics are from the category of edge-based QA methods. This subject proves the importance of edge-based features in exploring the spatial quality of pan-sharpened images. Moreover, Standard Deviation, Mean Gradient, and Spatial Frequency metrics can be considered as the weakest in detecting spatial distortion. Among indices without a reference image, JNB metric is showing more sensitivity to spatial distortions. The outstanding point is that three out of the top five best metrics are extracted from the image-processing domain and have never been used in the area of fusion. It can be concluded that image processing achievements in the field of image quality can be used to measure the quality of the pan-sharpened images as valuable quality metrics.

Due to the lack of robust and reliable spatial assessment methods in the area of fusion, this research has examined and compared several useful methods from image quality domains with existing metrics, so the best-proposed methods in this study can be generalized to many applications of pan-sharpened images.

# REFERENCES

Alimuddin, I., Sumantyo, J. T. S., & Kuze, H., 2012. Assessment of pan-sharpening methods applied to image fusion of remotely sensed multi-band data. International Journal of Applied Earth Observation and Geoinformation, 18, 165-175.

Alparone, L., Wald, L., Chanussot, J., Thomas, C., Gamba, P., & Bruce, L. M., 2007. Comparison of pansharpening algorithms: Outcome of the 2006 GRS-S data fusion contest. IEEE Transactions on Geoscience and Remote Sensing, 45(10), 3012-3021.

Amolins, K., Zhang, Y., & Dare, P., 2007. Wavelet based image fusion techniques—An introduction, review and comparison. ISPRS Journal of Photogrammetry and Remote Sensing, 62(4), 249-263.

Al-Wassai, F. A., & Kalyankar, D. N., 2012. A Novel Metric Approach Evaluation For The Spatial Enhancement Of Pan-Sharpened Images. arXiv preprint arXiv:1207.5064.

DadrasJavan, F., Samadzadegan, F., & Fathollahi, F., 2018. Spectral and Spatial Quality assessment of IHS and Wavelet Based Pan-sharpening Techniques for High Resolution Satellite Imagery. Advances in Image and Video Processing, 6(2), 01.

Ehlers, M., Klonus, S., Johan Åstrand, P., & Rosso, P., 2010. Multi-sensor image fusion for pansharpening in remote sensing. International Journal of Image and Data Fusion, 1(1), 25-45.

Ferzli, R., & Karam, L. J., 2009. A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB). IEEE transactions on image processing, 18(4), 717-728.

Garzelli, A., Nencini, F., & Capobianco, L., 2008. Optimal MMSE pan sharpening of very high resolution multispectral images. IEEE Transactions on Geoscience and Remote Sensing, 46(1), 228-236.

Jagalingam, P., & Hegde, A. V., 2015. A review of quality metrics for fused image. Aquatic Procedia, 4, 133-142.

Javan, F., Samadzadegan, F., & Reinartz, P., 2013. Spatial quality assessment of pan-sharpened high resolution satellite imagery based on an automatically estimated edge based metric. Remote Sensing, 5(12), 6539-6559.

Kang, X., Li, S., & Benediktsson, J. A., 2013, July. Pansharpening of remote sensing images with a matting model. In 2013 IEEE International Geoscience and Remote Sensing Symposium-IGARSS (pp. 1226-1229). IEEE.

Khan, M. M., Alparone, L., & Chanussot, J., 2008, July. Pansharpening quality assessment using modulation transfer function filters. In Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International (Vol. 5, pp. V-61). IEEE.

Mandhare, R. A., Upadhyay, P., & Gupta, S., 2013. Pixel-level image fusion using brovey transforme and wavelet transform. International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, 2(6), 2690-2695.

Naidu, V. P. S., 2010. Discrete cosine transform-based image fusion. Defence Science Journal, 60(1), 48-54.

Ni, Z., Ma, L., Zeng, H., Chen, J., Cai, C., & Ma, K. K., 2017. ESIM: Edge similarity for screen content image quality assessment. IEEE Transactions on Image Processing, 26(10), 4818-4831.

Nikolakopoulos, K., & Oikonomidis, D., 2015. Quality assessment of ten fusion techniques applied on Worldview-2. European Journal of Remote Sensing, 48(1), 141-167.

Palubinskas, G., 2015. Joint quality measure for evaluation of pansharpening accuracy. Remote Sensing, 7(7), 9292-9310.

Pálsson, F., 2013. Pansharpening and Classification of Pansharpened Images. Sc. A Thesis in Electrical and Computer Engineering at the, University of Iceland.

Qi, H., Jiao, S., Lin, W., Tang, L., & Shen, W., 2014. Contentbased image quality assessment using semantic information and luminance differences. Electronics Letters, 50(20), 1435-1436.

Ryan, R., Baldridge, B., Schowengerdt, R. A., Choi, T., Helder, D. L., & Blonski, S., 2003. IKONOS spatial resolution and image interpretability characterization. Remote sensing of environment, 88(1-2), 37-52.

Snehmani, Gore, A., Ganju, A., Kumar, S., & Srivastava, P. K., 2017. A comparative analysis of pansharpening techniques on QuickBird and WorldView-3 images. Geocarto International, 32(11), 1268-1284.

Thomas, C., & Wald, L., 2004, May. Assessment of the quality of fused products. In 24<sup>th</sup> EARSeL Symposium "New Strategies for European Remote Sensing" (pp. 317-325). Millpress.

Vijayaraj, V., O'Hara, C. G., & Younan, N. H., 2004, September. Pansharpening and image quality interface. In IGARSS 2004. 2004 IEEE International Geoscience and Remote Sensing Symposium (Vol. 4, pp. 2558-2560). IEEE.

Vivone, G., Alparone, L., Chanussot, J., Dalla Mura, M., Garzelli, A., Licciardi, G. A., ... & Wald, L., 2015. A critical comparison among pansharpening algorithms. IEEE Transactions on Geoscience and Remote Sensing, 53(5), 2565-2586.

Wald L, Ranchin T, Mangolini M., 1997. Fusion of satellite images of different spatial resolutions: assessing the quality of resulting images. Photogramm Eng Remote Sens. 63:691–699.

Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P., 2004. Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing, 13(4), 600-612.

Xue, W., Zhang, L., Mou, X., & Bovik, A. C., 2013. Gradient magnitude similarity deviation: A highly efficient perceptual image quality index. IEEE Transactions on Image Processing, 23(2), 684-695.

Yu, S., Zhang, W., Wu, S., Li, X., & Xie, Y., 2014, October. Applications of edge preservation ratio in image processing. In 2014 12th International Conference on Signal Processing (ICSP) (pp. 698-702). IEEE.

Zhang, Y., & Hong, G., 2005. An IHS and wavelet integrated approach to improve pan-sharpening visual quality of natural colour IKONOS and QuickBird images. Information Fusion, 6(3), 225-234.

Zhou, J., Civco, D. L., & Silander, J. A., 1998. A wavelet transform method to merge Landsat TM and SPOT panchromatic data. International journal of remote sensing, 19(4), 743-757.