Wald L., 2000. Quality of high resolution synthesised images: Is there a simple criterion ?. In Proceedings of the third conference "Fusion of Earth data: merging point measurements, raster maps and remotely sensed images", Sophia Antipolis, France, January 26-28, 2000, Thierry Ranchin and Lucien Wald Editors, published by SEE/URISCA, Nice, France, pp 99-103.

QUALITY OF HIGH RESOLUTION SYNTHESISED IMAGES: IS THERE A SIMPLE CRITERION?

Lucien Wald

Groupe Télédétection & Modélisation, Ecole des Mines de Paris,
BP 207, 06904 Sophia Antipolis cedex, France.
Phone: +33 (0) 4 93 95 74 49 - Fax: +33 (0) 4 93 95 75 35
E-mail: lucien.wald@cenerg.cma.fr

ABSTRACT

Methods exist which are synthesizing high spatial resolution high spectral content images from a set of low spatial resolution high spectral content images and high spatial resolution low spectral content images. There is a need to quantify the quality of these synthesized images for both the producers and customers of such fused products. Some protocols are discussed. The main concern is on the definition of a parameter characterizing the quality. The requirements on this parameter are discussed. A review of published parameters is made and a new one is proposed. From published works, it is shown that a good quality is achieved when the parameter is less than 3.

1. INTRODUCTION

In various applications, the benefit of obtaining multispectral images with the highest spatial resolution available has been demonstrated, particularly for vegetation, land-use, precision farming and urban studies. On the one hand, the high spatial resolution is necessary for an accurate description of the shapes, features and structures. On the other hand, depending on the application and the level of land cover complexity, the different types of land-use are better classified if high spectral resolution images are used. Hence, there is a desire to combine the high spatial and the high spectral resolutions with the aim of obtaining the most complete and accurate (in terms of spectral band) description of the observed area.

Several approaches of sensor fusion exist which apply on a data set comprising multispectral images B_{il} at a low spatial resolution l and images A_h at a higher spatial resolution h but with a lower spectral content. Examples of such a data set are the SPOT-XS (3 bands, 20 m) and SPOT-P (panchromatic, 10 m) images. These methods aim at constructing synthetic multispectral images B^*_{ih} having the highest spatial resolution available within the data set (e.g. the 3 XS bands at 10 m in the case of SPOT 1-3) which are close to reality by performing a high-quality transformation of the multispectral content when increasing the spatial resolution.

Examples of such approaches are

- Projection of original datasets into another space, substitution of one vector by the high resolution image and inverse projection into the original space. The common methods are IHS (Intensity, Hue, Saturation) method and the PCA (Principal Component Analysis) method (Carper *et al.*

1990). Some commercial softwares for the processing of images from e.g., Adobe or JASC companies, propose a function called transparency, which acts similarly to the IHS method.

- Relative spectral contribution. Here the spectral bands B_{il} are resampled at the highest resolution h; the results are called B'_{ih}. Then the synthesized band B^*_{ih} is equal in each pixel to B'_{ih} multiplied by a certain weight, depending upon the pixel. This weight represents the spectral contribution of the spectral band i to the panchromatic band. It is equal to the ratio of the panchromatic band and of the sum of the spectral bands B'_{ih} that are encompassed by the panchromatic band. Examples are the Brovey transform, the CNES P+XS method (Anonymous, 1986) and the generalized method (Wiemker *et al.*, 1998).
- Scale by scale description of the information content of both images and synthesis of the high-frequency information missing to transform the low spatial resolution images into high spatial resolution high spectral content images. The ARSIS concept (Ranchin, Wald, 1999) has developed in several methods (see a review in Ranchin, Wald), including the HPF method (Chavez *et al.* 1991).

Producers, *i.e.* providers of fused products, and customers, *i.e.* users of such fused products, may hesitate to select one of these methods or fused products. Commercial softwares often propose several different methods and it is not obvious for non -specialists to select one method or another for a given case. It follows that usually producers often use methods, which are not the most suitable for their customers.

Several comparisons between methods have been published and are regularly published. However results poorly disseminate in the community and there is lack of knowledge among producers regarding these methods, their advantages and limits. The lack of standardization of protocols for comparison does not add to the clarity of the results. Some efforts have been made recently (Wald *et al.* 1997) but a lot still remain.

2. A PROTOCOL FOR COMPARISON OF METHODS

The merging methods under concern aim at constructing synthetic images B^*_{ih} close to the reality. Wald *et al.* (1997) established the properties of such synthetic images:

- Any synthetic image B^*_{ih} once degraded to its original resolution l, should be as identical as possible to the original image B_{il}.
- Any synthetic image B^*_{ih} should be as identical as possible to the image B_{ih} that the corresponding sensor would observe with the highest spatial resolution h.
- The multispectral set of synthetic images B^*_{ih} should be as identical as possible to the multispectral set of images B_{ih} that the corresponding sensor would observe with the highest spatial resolution h.

A critical review of the protocols found in the literature has been made. It was found that the protocol proposed by Wald *et al.* (1997), distinguished by the American Society for Photogrammetry and Remote Sensing, is presently that permitting the most complete checking of the three properties. A similar approach, but incomplete, was proposed earlier by Mangolini *et al.* (1992, 1995) and Munechika *et al.* (1993). One of the most common protocols consists in interpolating low resolution data up to the high resolution, and then making the difference pixel per pixel (Carper *et al.* 1990). In any case, are the interpolated images representatives of what should be observed by a similar sensor with a higher resolution, and these interpolated images cannot constitute a valid reference. It follows that this protocol is not valid and should not be used. Other protocols compare some statistical quantities derived from the original dataset and from the synthesized images. However, Wald *et al.* discussed the non-preservation of statistical distribution when changing spatial resolution in the light of several published works. They concluded that any protocol based upon the comparison of statistical quantities (e.g., histogram) is not valid.

It is not our purpose to discuss any further the protocol. This discussion shows that presently an accepted protocol has been worked out. It is simple to implement and it may become the standard approach agreed upon by all the producers of fused products which scopes are in the frame of this discussion.
Wald *et al.* have underlined the importance of the landscape upon the quality of the fusion. They recommend testing the quality on very heterogeneous landscape such as urban areas, which offer a large variety of spectral signatures as well as of space scales of non-fractal nature.

3. THE NEED FOR A SINGLE QUANTITY DESCRIBING THE QUALITY

Even if a satisfying protocol is adopted, there is a further need for a simple characterization of the quality of the product of the fusion process, which can be associated to each product and qualifies it. It would greatly help producers to select methods and improve their production lines, and customers to make their choice among products and to assess the impact of this quality on further processing.
The protocol of Wald *et al.* is based upon the differences between the synthesized images and the actual ones. These differences are summarized by various statistical quantities, which characterize the performance in synthesizing an image in a given spectral band, the spectral signature, and the actual *n*-tuplets (and particularly those which are predominant in multispectral classification). Other published works also use statistical quantities. Root mean square errors $RMSE(B_i)$ are very often used. On the contrary biases and mean values are given seldom.
These quantities are very useful to fully understand the performances of a method. However, our experience shows that there are too many figures. There is a need for a quantity, which gives a quick insight of the quality. What we are looking for, is a number simple to understand which is a good indicator of the overall error of the fused product. The closer to 0 this number, the better the product.
This quantity should fill three requirements:
First requirement. It should be independent of units, and accordingly of calibration coefficients and instrument gain. Customers fairly seldom take calibration coefficients into account. Some fusion methods can be applied to unitless quantities or to radiances. Consequently the quality parameter should be independent of units.
Second requirement. This quantity should be independent of the number of spectral bands under consideration. This is a *sine qua non* condition to compare results obtained in various conditions. For example, a method applied to the SPOT-XS 2 and -XS 3 should produce similar quality than when applied to the three bands.
Third requirement. This quantity should be independent of the ratio of the scales *h/l*. This permits to compare results obtained in different cases, like the fusion of Landsat-TM and SPOT-P, and the fusion of SPOT-XS with SPOT-P.
Munechika *et al.* (1993) propose to use the following quantity to globally characterize the quality of the fused product. They called it total error.

$$\text{Total error} = \sum_{i=1}^{N} RMSE(B_i)$$

This total error does not obey any of the three requirements. In particular it is sensitive to the changes from numerical counts to radiances. Ranchin, Wald (1999) generalize this formula in order to be able to compare errors obtained from different methods, different cases and different sensors. Let M_i be the mean value for the original spectral image B_i. Let M be the mean radiance of the N images B_i:

$$M = (1/N) \sum_{i=1}^{N} M_i$$

The relative average spectral error *RASE* is expressed in percent and characterizes the average performance of a method in the considered spectral bands:

$$RASE = \frac{100}{M}\sqrt{\frac{1}{N}\sum_{i=1}^{N} RMSE(B_i)^2}$$

The RASE obeys the first and second requirements, but not the third one.
Let M_i be the mean value for the spectral image B_i. We propose the following quantity

$$New\ quantity = 100\ \frac{h}{l}\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[RMSE(B_i)^2 \big/ (M_i)^2 \right]}$$

It is more robust than RASE with respect to calibration and changes of units. It obeys the second requirement and the ratio *h/l* is taking into account the various resolutions.

4. COMPARISON BETWEEN THE VARIOUS QUALITY QUANTITIES

These various quantities are now computed for several cases found in the literature. Most of them are dealing with the SPOT case using different methods, but other cases are available. For each case, we have reported the comments made by the authors on the overall quality of the fused products (Table 1). In all cases, the protocol was that of Wald *et al.* (1997). The cases used by Mangolini *et al.* (1992, 1995) have been re-processed by Wald *et al.* (1997) or by Ranchin, Wald (1999). The case given by Munechika *et al.* (1993) cannot be fully used since they do not provide means of original images, nor biases. This remark also holds for most of the published works, which do not provide enough quantitative information to compute one of the three quantities. This situation enforces the comments made on the lack of clarity on the state-of-the-art.

Case: method, area, sensor (ref.)	Low res.	High res.	Comments	Total error	RASE	New quantity
Brovey. SPOT XS and P. Barcelona (Ranchin, Wald, 1999)	40 m	20 m	bad quality	105.2	65.5	32.7
IHS. SPOT XS and P. Barcelona (Ranchin, Wald, 1999)	40 m	20 m	bad quality	21.6	13.5	6.7
PCA. SPOT XS and P. Barcelona (Ranchin, Wald, 1999)	40 m	20 m	bad quality	13.6	8.6	4.5
Duplication. SPOT XS and P. Barcelona (Ranchin, Wald, 1999)	40 m	20 m	bad quality	12.2	7.6	3.9
CNES P+XS. SPOT XS and P. Barcelona (Wald *et al*. 1997)	40 m	20 m	bad quality	10.7	6.7	3.3
ARSIS model 1. SPOT XS and P. Barcelona (Ranchin, Wald, 1999)	40 m	20 m	good quality	9.3	6.1	3.0
ARSIS model 2. SPOT XS and P. Barcelona (Ranchin, Wald, 1999)	40 m	20 m	good quality	8.1	5.2	2.6
ARSIS model RWM. SPOT XS and P. Barcelona (Wald *et al*. 1997)	40 m	20 m	good quality	6.5	4.1	2.1
ARSIS model RWM. SPOT XS and P. Barcelona (Wald *et al*. 1997)	80 m	40 m	good quality	8.1	5.2	2.6
Arsis model RWM. SPOT-4 (4 bands) and -P. Baja (Hungary) (unpublished)	40 m	20 m	good quality	6.8	7.6	2.3
Bicubic interpolation of Landsat TM-1, -2, -5 (Blanc *et al*. 1998)	90 m	30 m	bad quality	1.5	15.3	5.2
Arsis 1/3#1. Landsat TM-1, -2, -5 and SPOT-P (Blanc *et al*. 1998)	90 m	30 m	bad quality	1.4	14.0	5.0

Arsis 1/3#3. Landsat TM-1, -2, -5 and SPOT-P (Blanc *et al*. 1998)	90 m	30 m	bad quality	1.4	13.5	4.6
Arsis model RWM. Landsat TM-6 and TM-4. Nantes (Wald, Baleynaud 1998)	240 m	120 m	good quality	3.0	2.0	1.0

Table 1. Comparison between the three quality parameters.

Two cases are not fusion but only duplication or bicubic interpolation of the multispectral images, which do not call at all on the high resolution image. They are mentioned here to help in assessing the overall performances of fusion methods. Six cases only are reported as being of good quality by the authors: the cells in the table are in light grey.

This Table does not intend to compare methods. It shows how the three quality parameters react to changes in case conditions.

The total error of Munechika *et al.* (1993) decreases as the RMSE for each band decreases. It is very sensitive to the changes in the number of spectral bands. The last case with only one band (TM-6) presents a total error of 3.0 while for SPOT it is closer to 10. It is also sensitive to the changes in units. For the Landsat cases with TM-1, -2 and -5 fused with SPOT-P or not, the total error is about 1.5, that is less than for the last case with only TM-6. But these cases are not of satisfying quality while the last one is. Accordingly the total error cannot represent in a simple way the overall quality in all the cases shown here.

The relative average spectral error *RASE* behaves better. It offers a better tendency to decrease as the quality increases. It is fairly independent of units provided they are the same for all bands. It is also fairly independent of the number of bands provided the range of values for each band is fairly constant. In the case of SPOT-4 and its four bands, the *RASE* has the value of 7.6 for a good quality, while for SPOT-2 (cases of Barcelona) such a value denotes a bad quality.

The new quantity exhibits a strong tendency to decrease as the quality increases. It behaves correctly whatever the number of bands is because it uses for each band the rmse relative to the mean of the band. This definition makes also this quantity independent of the calibration or changes in units, allowing even changes from band to band. One may notice that "good quality" always corresponds to values less than 3, in any of the cases displayed in the Table.

5. CONCLUSION

A new quantity has been proposed which aims at providing a quick but accurate insight of the overall quality of a fused product. Using several different cases found in the literature, it has been found that this quantity behaves better than those already proposed. It has also been found that this quantity reflects the conclusions of the different authors relative to the methods. Accordingly it may serve to broadly assess the quality of a method. On the other hand, very similar values of this quality parameter are found for different cases, which have been declared satisfactory by their authors. A threshold of satisfaction may be set to 3 for a product. Below 3, the error is small and the product is of good quality. Above 3, the error is large and the product is of lower quality. The quality decreases as the error increases.

To conclude, this preliminary study is very encouraging. We found that

- a protocol is now available
- a simple and robust quality parameter can be defined
- this parameter qualifies the method
- it may also qualify any fused product, though more cases should be examined.

These results indicate that it should be possible in a near future that producers of fused products deliver a standardized assessment of the quality of their products. This would allow them to better design and improve their production chains, and would allow customers to better select the products and improve their efficiency.

ACKNOWLEDGMENTS

This work was partly supported by SPOT-Image, Centre National d'Etudes Spatiales and Eurimage, which provide images from SPOT and Landsat.

REFERENCES

Anonymous, 1986. *Guide des utilisateurs de données SPOT,* 3 tomes, Editeurs CNES et SPOT-Image, Toulouse, France.

Blanc, Ph., T. Blu, T. Ranchin, L. Wald, and R. Aloisi, 1998. Using iterated rational filter banks within the ARSIS concept for producing 10 m Landsat multispectral images. *International Journal of Remote Sensing*, 19(12):2,331-2,343.

Carper, W. J., T. M. Lillesand, and R. W. Kiefer, 1990. The use of Intensity Hue Saturation transformations for merging SPOT panchromatic and multispectral image data. *Photogrammetric Engineering & Remote Sensing*, 56(4):459-467.

Chavez, P. S. Jr., S. C. Sides, and J. A. Anderson, 1991. Comparison of three different methods to merge multiresolution and multispectral data: Landsat TM and SPOT panchromatic. *Photogrammetric Engineering & Remote Sensing*, 57(3):265-303.

Mangolini, M., T. Ranchin, and L. Wald, 1992. Procédé et dispositif pour l'amélioration de la résolution spatiale d'images à partir d'autres images de meilleure résolution spatiale. French patent n° 92-13961, 20 novembre 1992.

Mangolini, M., T. Ranchin, and L. Wald, 1995. Evaluation de la qualité des images multispectrales à haute résolution spatiale dérivées de SPOT. *Bulletin Société Française de Photogrammétrie et Télédétection*, 137, pp. 24-29.

Munechika, C. K., J. S. Warnick, C. Salvaggio, and J. R. Schott, 1993. Resolution enhancement of multispectral image data to improve classification accuracy. *Photogrammetric Engineering & Remote Sensing*, vol. 59, 1, pp. 67-72.

Ranchin T., and L. Wald, 1999. Fusion of high spatial and spectral resolution images: the ARSIS concept and its implementation. *Photogrammetric Engineering & Remote Sensing*, 63(6): 691-699.

Wald, L., and J. M. Baleynaud, 1999. Observing air quality over the city of Nantes by the means of Landsat thermal IR data. *International Journal of Remote Sensing,* 20, 5, 947-959.

Wald, L., T. Ranchin, and M. Mangolini, 1997. Fusion of satellite images of different spatial resolutions: assessing the quality of resulting images. *Photogrammetric Engineering & Remote Sensing*, 63(6): 691-699.

Wiemker, R., B. Prinz, G. Meister, R. Franck, and H. Spitzer, 1998. Accuracy assessment of vegetation monitoring with high spatial resolution satellite imagery. *International Archives of Photogrammetry and Remote Sensing*, vol. XXXII, Part 7, 285-292.