

Analysis of Changes in Quality Assessment with Scale

Claire Thomas

Ecole des Mines de Paris,
BP 207, 06904 Sophia Antipolis cedex, France
Claire.thomas@ensmp.fr

Lucien Wald

Ecole des Mines de Paris,
BP 207, 06904 Sophia Antipolis cedex, France
Lucien.wald@ensmp.fr

Abstract - Our work deals with the assessment of the quality of multimodal images synthesized at a better spatial resolution by the means of another image having such a resolution. In absence of reference images, the current protocols recommend to degrade spatially both sets of images and to perform fusion on these two new sets, thus producing synthesized images at the original low resolution. The quality budget is drawn at this scale and it is assumed that it is close to or worse than the one, which would have been drawn at high resolution if the reference images were available. This extrapolation hypothesis is the major limitation of the application of protocols. The validity of this hypothesis is studied for 44 case studies. Several distances are selected. They are compounded into several quality budgets, each of them giving an overall idea on the quality. We analyze changes in quality budgets with scales. It is found that the extrapolation hypothesis is validated in most cases. Nevertheless, the need for further investigation is underlined.

Keywords: quality budget, ERGAS, multiresolution analysis, extrapolation, image fusion, satellite images, synthesis of multi-modality images, pan-sharpened.

1 Introduction

Several approaches of sensor fusion exist which apply on a data set comprising multimodal images at a low spatial resolution l and images at a higher spatial resolution h but with a lower spectral content. These methods aim at constructing synthetic multimodal images having the highest spatial resolution h available within the data set [1]. Under concern in this paper are only those methods that claim to provide synthetic images close to reality when enhancing the spatial resolution, and not those that provide only a better visual representation of the data set.

More specifically, this paper deals with the assessment of the quality of the products resulting from the application of one of these methods, the so-called fused products. Several protocols exist [2, 3, 4]. They all face the problem of the lack of multimodal reference at high resolution h against which the fused products may be compared. These three protocols suggest alleviating this lack of reference by moving the problem in the scales. The multimodal images and the high spatial resolution image are each degraded to lower resolutions,

so that the high resolution image has now the original resolution l of the multimodal images. Then the fusion is performed on these two new sets of data. It results into the synthesis of images at the original resolution l of multimodal images. At this scale, the problem of reference does not exist since if the fusion process is perfectly made, one should retrieve the original multimodal images.

The quality assessment is then drawn at this scale l and it is assumed that it is close to or worse than the one that would have been drawn at high resolution h if the reference images were available. This extrapolation hypothesis is the major limitation of the application of protocols for quality assessment.

The validity of this hypothesis is evaluated in a heuristic approach by the means of 44 case studies. After a brief review of widely-used distances, we select some of them and justify our choices. These distances are gathered into several quality budgets, each of them being able to qualify fused products. For each case study, we analyze changes in quality budgets with scales and we conclude on the validity of the hypothesis.

The work is performed by the means of the platform for quality assessment developed by [5], exploiting the protocol of [4]. The degradation in resolution was made by the means of the UWT (undecimated wavelet transform) called "à trous" (with holes) [6]. The fusion method selected here is the well-known method UWT-M2 proposed by [7, 8].

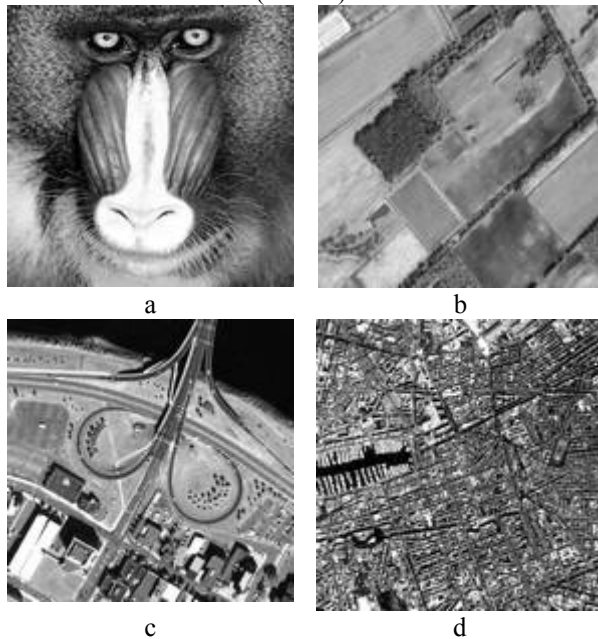
2 The case studies

We have selected the well-known images of the mandrill and Lenna. These images are made of three modalities: red, green, and blue, at the resolution h . The red modality was arbitrarily selected as the high resolution image and the two others were degraded to the resolution $l=2h$, or $l'=4h$.

The other images were acquired by satellite observing systems: SPOT-2 ($h=10\text{ m}$, $l=20\text{ m}$), SPOT-5 ($h=5\text{ m}$, $l=10\text{ m}$), Ikonos ($h=1\text{ m}$, $l=4\text{ m}$) and Quickbird ($h=0.7\text{ m}$, $l=2.8\text{ m}$). These satellite images cover 14 geographical sites with a ratio 2 (Table 1) and 26 sites with a ratio 4 (Table 2). The type of landscape ranges from urban areas to agricultural areas. Their common characteristics are the large amount of high frequencies (small details) and different multimodal vectors present

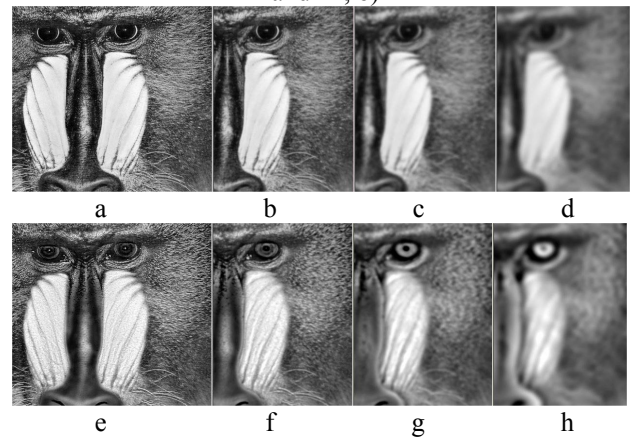
in each image. Such images were recommended by [3] to enhance the ability of a fusion method to synthesize features of small size on the one hand and multimodal vectors on the other hand. A sample of the image diversity is presented figure 1.

Figure 1. image diversity : extreme cases. a) mandrill, b) Ikonos Hasselt (Belgique) fields, c) Quickbird Fredrickton (Canada) interchange and d) SPOT5 Marseille (France) downtown.



has a resolution of $res1=2$, c : $res2=4$, and finally d : $res3=8$.

Figure 2. image diversity : extreme cases. a) mandrill, b)



Images from e to h correspond to fused blue modalities at the same resolutions than respectively a to d images. A quality budget is drawn between images d and h , then on c and g , and so on. At last, we study the evolution of these quality budgets along scales.

Tables 1 and 2 report a summary of the different case studies. They specify the satellite and the geographical area, the number of extracts selected in this area, the original spatial resolution of multispectral images and the resolutions of the successive synthesized images.

For each case, all modalities (high and low resolution) were degraded by a factor equal to the ratio (l/h) and the fusion is performed at these new resolution. For instance, Quickbird multispectral modalities are degraded from their original low resolution $res1=2.8$ m down to $res2=11.2$ m and in a similar way, the panchromatic high resolution image is degraded from the resolution $res0=0.7$ m down to $res1$. Fusion at these new resolutions stems into fused products at $res1=2.8$ m, were references are available, which are original multispectral images. The same operation is then performed onto these panchromatic at 2.8 m and these multispectral at 11.2 m, creating respectively images at $res2=11.2$ m and $res3=44.8$ m. then down to $res3=44.8$ m. In addition when the ratio is 4, we created images at an intermediate resolution $res_{int}=5.6$ m. Though of no direct concern, such images increases the number of available images and consequently helped in analysis of results.

Figure 2 show with the example of the blue modality of mandrill how a fused product is created and compared to the reference multimodal image at each resolution, by the mean of quality budgets, presented in section 3. Images from a to d are the successive coarser approximations of the blue modality. They are separated by a ratio 2, simulated thanks to UWT "à trous" algorithm. For Lenna and mandrill, we can't talk about real spatial resolution, but more about relative ones compared to the original color image resolution. Here, a is original blue modality $res0=1$. Then, successively b

Table 1. List of images selected for ratio 2

| Image source | Number of extracts | Original resolution n | Constructed resolutions |
|--------------------------|--------------------|-------------------------|--------------------------|
| Mandrill and Lenna | 1 for each | 1 | 1, 2, 4 (relative value) |
| SPOT-2, Barcelona, Spain | 1 | 20 m | 20, 40, 80 m |
| SPOT5, Marseille, France | 7 | 10 m | 10, 20, 40 m |
| SPOT5, Toulouse, France | 6 | 10 m | 10, 20, 40 m |

Table 2. List of images selected for ratio 4

| Image source | Number of extracts | Original resolution n | Constructed resolutions |
|---------------------------|--------------------|-------------------------|--------------------------|
| Mandrill and Lenna | 1 for each | 1 | 1, 2, 4 (relative value) |
| Ikonos, Hasselt, Belgium, | 8 | 4 m | 4, 8 and sometimes 16 m |

| | | | |
|--------------------------------------|---|-------|-------------------------------------|
| Ikonos, Fredrickton, Canada | 5 | 4 m | 4, 8 and sometimes 16 m |
| Quickbird, Madrid, Spain | 6 | 2.8 m | 2.8, 5.6, 11.2 m |
| Quickbird, Fredrickton, Canada | 7 | 2.8 m | 2.8, 5.6 and sometimes 11.2 m |

3 Distances and quality budgets

A distance quantifies the discrepancy between a reference and the fused product. A quality budget is a composition of one or more distances. The distances (respectively quality budget) are of two types: monomodal and multimodal. A monomodal distance (respectively quality budget) applies to a single modality, while a multimodal distance (respectively quality budget) applies to several modalities. Several distances are well-known such as the correlation coefficient; others have been recently published and we propose new ones. These distances are implemented in the platform [5]. According to the published results and our own knowledge, we have selected the most discriminant distances for our study.

The monomodal distances selected are:

- the difference in variance relative to the variance of the reference, i.e., the variance of the fused product minus the variance of the reference for each modality, divided by the variance of the reference, *diffVarRel* [3]. For a perfect fusion process, the ideal value should be 0;
- the standard-deviation of the image of the difference made pixel per pixel between the fused and reference images [2, 3]. This standard-deviation is expressed relative to the mean value of the reference image, σ_{rel} . The ideal value is 0;
- the correlation coefficient between the fused and reference images, *cc*. The ideal value is 1;
- the quality index *Q* [9]. If *M* and σ denote respectively the mean and the standard-deviation of the reference image, and if *M** and σ^* denote the same quantities for the fused image, this index *Q* is given by:

$$Q = cc [2 M M^* / (M^2 + M^{*2})] [2 \sigma \sigma^* / (\sigma + \sigma^*)] \quad (1)$$

The ideal value is 1.

The bias, i.e., the difference between *M* and *M** was recommended by [3, 4]. This distance is not used here because it is equal to zero in the case of our fusion method: it is not discriminant.

The multimodal distances selected are:

- the ERGAS [8, 10]. If B_k denotes the image for modality *k*, we note the mean value as M_k . $RMSE(B_k)$ denotes the root mean square difference between the reference and fused images for modality *k*.

If *N* is the number of modalities, the ERGAS is given by:

$$ERGAS = 100 \frac{h}{l} \sqrt{\frac{1}{N} \sum_{k=1}^N \frac{(RMSE(B_k))^2}{M_k^2}} \quad (2)$$

The ideal value is 0;

- the SAM [11]. If V_{Bk} and V_{Bk^*} denote the multimodal reference and fused vectors at each pixel *i* and *M* is the number of pixels, the SAM (spectral angle mapper) is:

$$SAM(V_{Bk}, V_{Bk^*}) = \frac{1}{M} \sum_{i=1}^M \arccos \left(\frac{\langle V_{Bk}, V_{Bk^*} \rangle}{\|V_{Bk}\| \|V_{Bk^*}\|} \right) \quad (3)$$

SAM is expressed in degrees. The ideal value is 0;

- the mean difference of norms $\|V_{Bk}\| - \|V_{Bk^*}\|$, relative to the mean of the norms of the reference image, *biasRelNorm*. The ideal value is 0;
- the standard-deviation of the image of the difference of norms, relative to the mean of the norms of the reference image, *σRelNorm*. The ideal value is 0;
- the mean of the norm of the resulting vector $(V_{Bk} - V_{Bk^*})$ for each pixel and the standard-deviation, respectively $V_{res} moy$ and $V_{res} \sigma$. In both cases, the ideal value is 0.

As for the quality budget, we selected for monomodal budget:

- the correlation coefficient *cc*;
- the quality index *Q*;
- the vector (σ_{rel}, cc) ;
- the vector $(\sigma_{rel}, cc, diffVarRel)$;

The index *Q* is easily computed from *cc*. Comparing *cc* and *Q* permits to assess the benefit of *Q*. For multimodal budget, we selected the following quantities:

- the SAM;
- the vector made of $V_{res} moy$ and $V_{res} \sigma$.

Finally, we selected two quantities to assess the budget on a global basis:

- the ERGAS;
- the vector made of *Q* and *SAM*.

The analysis of the extrapolation hypothesis consists in building quality budgets at all scales and then, studying the discrepancies between the quality budgets for two successive scales. For each image, we have two fused products, one at *res1* and one at *res2*. For each resolution, a quality budget is drawn and the differences between the values obtained at both resolutions are compared. If the quality budget is better at *res1* than at *res2*, the hypothesis is validated. Otherwise, it is not.

Actually, we studied two hypotheses. The first one “the quality budget is better at *res1* than at *res2*”. Let note *dist(res1)* or *dist(res2)* any distance at resolution *res1* or *res2*. The first hypothesis is validated if within a given quality budget, each distance obeys the following:

$$dist(res1) \leq dist(res2) \quad (4)$$

except for distances cc and Q , for which the condition is:

$$\text{dist}(res1) \geq \text{dist}(res2) \quad (5)$$

The second case is less stringent: the quality budget is better at $res1$ than at $res2$ or close to that at $res2$. This second hypothesis is validated if within a given quality budget, each distance obeys the following:

$$\text{dist}(res1) \leq \text{dist}(res2) + \delta \quad (6)$$

except for distances cc and Q , for which the condition is:

$$\text{dist}(res1) \geq \text{dist}(res2) - \delta \quad (7)$$

The quantity δ is given in Table 3. It was set up from published results and our own experience.

Table 3. Interval δ allowed to each distance

| <i>Distance</i> | <i>Interval</i> | <i>Distance</i> | <i>Interval</i> |
|-------------------|-----------------|---|-----------------|
| <i>diffVarRel</i> | 2.5% | <i>SAM</i> | 0.5° |
| σ_{rel} | 2.5% | <i>biasRelNorm</i> | 0.05% |
| <i>cc</i> | 0.025 | <i>σRelNorm</i> | 2.5% |
| <i>Q</i> | 0.025 | <i>V_{res}moy</i> | 2.5 |
| <i>ERGAS</i> | 0.5 | <i>V_{res}σ</i> | 2.5 |

4 Results

The results depend upon the quality budget under concern. Almost identical results are obtained with the quality budgets cc and Q , as expected.

Table 4 reports on the percentage of cases respecting the first hypothesis for the 7 quality budgets and ratios 2 and 4. What appears first is that the results strongly depend upon the budget and the ratio. For several budgets, the first hypothesis is not often validated (percentage less than 50%) while for others this percentage is much larger (greater than 70%). As a whole, the first hypothesis is more often validated for ratio 4 than for ratio 2, except for quality budget *SAM*.

Table 4. Percentage of cases respecting the first hypothesis.

| <i>Quality budget</i> | <i>Ratio 2</i> | <i>Ratio 4</i> |
|--|----------------|----------------|
| <i>cc</i> or <i>Q</i> (monomodal) | 55% | 88% |
| (σ_{rel}, cc) (monomodal) | 19% | 58% |
| $(\sigma_{rel}, cc, diffVarRel)$ (monomodal) | 19% | 58% |
| <i>SAM</i> (multimodal) | 81% | 71% |
| $(V_{res}moy, V_{res}\sigma)$ (multimodal) | 38% | 41% |
| (Q, SAM) (global) | 53% | 69% |
| <i>ERGAS</i> (global) | 44% | 65% |

From this table, one may conclude that the first hypothesis is

- very often validated for quality budget *SAM* whatever the ratio. The percentage is 70-80%;
- very often validated for quality budget *cc* or *Q* for ratio 4;

- quite often validated for other quality budgets for ratio 4, except $(V_{res}moy, V_{res}\sigma)$;
- not often validated for ratio 2 as a whole.

Table 5 reports on the percentage of cases respecting the second hypothesis. As before, the results depend upon the budget and the ratio but in a lesser extent. Of course, the percentages are greater than before because the second hypothesis is less stringent than the first one. For all budgets and ratios, the second hypothesis is almost always validated (percentage close to 90-100%) except for quality budget $(V_{res}moy, V_{res}\sigma)$ and ratio 4 (47%), quality budget *ERGAS* and ratio 2 (69%) and quality budget $(\sigma_{rel}, cc, diffVarRel)$ and ratio 4 (80%).

As a whole, the second hypothesis is more often validated for ratio 4 than for ratio 2, except for quality budget *SAM*.

Table 5. Percentage of cases respecting the second hypothesis

| <i>Quality budget</i> | <i>Ratio 2</i> | <i>Ratio 4</i> |
|--|----------------|----------------|
| <i>cc</i> or <i>Q</i> (monomodal) | 94% | 100% |
| (σ_{rel}, cc) (monomodal) | 87% | 92% |
| $(\sigma_{rel}, cc, diffVarRel)$ (monomodal) | 87% | 80% |
| <i>SAM</i> (multimodal) | 88% | 94% |
| $(V_{res}moy, V_{res}\sigma)$ (multimodal) | 88% | 47% |
| (Q, SAM) (global) | 91% | 97% |
| <i>ERGAS</i> (global) | 69% | 88% |

One may conclude that the second hypothesis is

- almost often validated for quality budgets *cc*, *Q*, (σ_{rel}, cc) , *SAM* and (Q, SAM) whatever the ratio. The percentage is 90-100%;
- very often validated for quality budgets $(\sigma_{rel}, cc, diffVarRel)$ and *ERGAS* for both ratios. The percentage is 70-80%;
- almost often validated for quality budget $(V_{res}moy, V_{res}\sigma)$ but for ratio 2 only;
- not often validated for quality budget $(V_{res}moy, V_{res}\sigma)$ for ratio 4.

5 Conclusions

This work shows that the protocols of [2, 3, 4] may be applied in many cases. The hypothesis “the quality assessment drawn at scale l is close to or worse than the one which would have been drawn at high resolution h if the reference images were available” has been validated in most cases in our study.

We found that when the images are noisy, such as the images of the sea where the ratio signal-to-noise is low, the hypothesis is not validated, with the exception of some quality budgets based on the correlation coefficient.

We also found that the verification of the hypothesis strongly depends upon the quality budget and the ratio.

Some questions remain and more investigation is still needed to conclude firmly on the validation of the hypothesis.

The influence of the fusion method on the results is unknown. *A priori*, the investigation method is free from this influence. Nevertheless, the quality of the fusion process has certainly an impact on the validation of the hypothesis.

The images used for this investigation were re-sampled. The influence of the re-sampling process should be studied. Of particular interest, is the non-orthogonality of the undecimated wavelet transform used to degrade the images. How this affects the results should be investigated.

We think also that the Modulation Transfer Function of the instrument may play a role in the validation of the hypothesis as it plays a role in the quality of the fused product.

References

- [1] L. Wald, *Data Fusion. Definitions and Architectures - Fusion of Images of Different Spatial Resolutions*, Presses de l'Ecole, Ecole des Mines de Paris, Paris, France, ISBN 2-911762-38-X, 200 p., 2002.
- [2] C. K. Munechika, J. S. Warnick, C. Salvaggio and J. R. Schott, Resolution enhancement of multispectral image data to improve classification accuracy. *Photogrammetric Engineering & Remote Sensing*, 59, 1, 67-72, 1993.
- [3] L. Wald, T. Ranchin and M. Mangolini, Fusion of satellite images of different spatial resolutions: assessing the quality of resulting images. *Photogrammetric Engineering & Remote Sensing*, 63, 6, 691-699, 1997.
- [4] C. Thomas and L. Wald, Assessment of the quality of fused products. In Proceedings of the 24th EARSeL Symposium "New Strategies for European Remote Sensing", 25-27 May 2004, Dubrovnik, Croatia, Oluic (ed.), Millpress, Rotterdam, ISBN 90 5966 003 X, pp. 317-325, 2005.
- [5] C. Thomas, T. Ranchin, L. Wald and J.-C. Laneri, A modular platform for fusion of images. In Proceedings of the Fourth International Conference Physics in Image and Signal Processing PSIP'2005, January 31-February 2 2005, Toulouse, France, ISBN 2-912328-22-5, pp. 269-273, 2005.
- [6] P. Dutilleul, An implementation of the "algorithme a trous" to compute the Wavelet Transform. In: Combes, J. M., Grossman, A., Tchamitchian, Ph. (Eds.), *Wavelets: Time-Frequency Methods and Phase Space*. Springer, Berlin, pp. 298-304, 1989.
- [7] T. Ranchin and L. Wald, Fusion of high spatial and spectral resolution images: the ARSIS concept and its implementation. *Photogrammetric Engineering and Remote Sensing*, 66(1), 49-61, 2000.
- [8] T. Ranchin, B. Aiazzi, L. Alparone, S. Baronti and L. Wald, Image fusion. The ARSIS concept and some successful implementation schemes. *ISPRS Journal of Photogrammetry & Remote Sensing*, 58, 4-18, 2003.
- [9] Z. Wang and A.C. Bovik, A universal image quality index. *IEEE Signal Processing Lett., Transactions on Pattern Analysis and Machine Intelligence*, 9, 81-84, 2002.
- [10] L. Wald, Quality of high resolution synthesized 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, 2000.
- [11] A. Garzelli, F. Nencini, L. Alparone, B. Aiazzi and S. Baronti, Pan-sharpening of multispectral images: A critical review and comparison. In Proceedings of IEEE Geoscience and Remote Sensing Symposium IGARSS'04, pp. 81-84, 2004.