Comparison and evaluation of quality criteria for hyperspectral imagery

Emmanuel Christophe[a,b], Dominique Léger[a] and Corinne Mailhes[b]

[a]ONERA DOTA, 2 avenue E. Belin BP 4025, 31055 Toulouse Cedex 4, FRANCE
[b] IRIT - TéSA - ENSEEIHT, 2 rue C. Camichel BP 7122, 31071 Toulouse Cedex 7, FRANCE

ABSTRACT

Hyperspectral data appears to be of a growing interest over the past few years. However, applications for hyperspectral data are still in their infancy. Handling the significant size of hyperspectral data presents a challenge for the user community. To enable efficient data compression without losing the potentiality of hyperspectral data, the notion of data quality is crucial for the development of applications. To assess the data quality, quality criteria relevent to end-user applications are required.

This paper proposes a method to evaluate quality criteria. The purpose is to provide quality criteria corresponding well to the impact of degradation on end-user applications. Several quality criteria adapted to hyperspectral context are evaluated. Finally, five criteria are selected to give a good representation of the degradation nature and level affecting hyperspectral data.

Keywords: hyperspectral, quality criteria, evaluation, compression, end-user applications.

1. INTRODUCTION

Airborne hyperspectral data have been available since the early 1980s, and their applications for geological, agricultural or military purposes are well established. Later, in 2000, spaceborne hyperspectral data became available upon the launch of Hyperion by NASA. Hyperspectral sensors collect the spectrum for each pixel of the image. Typically, hyperspectral images comprise of hundreds of narrow and contiguously bands from 0.4 to 2.5 micrometers. This fact results in an important amount of data produced by the sensor, thereby making the data compression a crucial step in the process of hyperspectral image acquisition and processing.

When dealing with lossy data compression, this is important to define a fidelity criterion or a distortion measure, able to quantify properly the information loss due to compression algorithms. Within the field of hyperspectral image lossy compression, most of papers evaluate the impact of the compression by using only classical distance measurements like MSE (Mean Square Error) or PMAD (Percentage Maximum Absolute Difference). However, it is widely known that even for ordinary images, standard metrics do not reflect the perceived information loss well. Therefore, few papers also use criteria adapted to hyperspectral.

In image or video processing, objective quality criteria are compared with subjective mean opinion scores from panels of observers. Indeed as the human viewer is the end user, he is important in the process of quality evaluation. However, hyperspectral data are generally not used directly by human viewers, due to the large amounts of data generated, but are instead processed by automatic algorithms. The development of objective quality criteria for hyperspectral data should then be done according to the impact of degradations on these algorithms. Given the sensitivity of application algorithms and the practical difficulties in setting them up (e.g. computational cost, need for ground truth, simulation of targets, or expert abilities etc.) it would not be viable to use them as direct quality measures. The purpose of evaluating quality criteria in general is to find the objective criteria corresponding well to the impact on the application or the viewer.

Further author information: (send correspondence to E. Christophe)
E. Christophe: Emmanuel.Christophe@tesa.prd.fr, Telephone: 33 5 61 58 80 15

The goal of this paper is to define quality criteria that are suitable for hyperspectral images. The first part presents hyperspectral images and their typical applications. The second part elaborates on quality criteria and some quality criteria are proposed. The third part defines a process of evaluation to compare performances of different quality criteria. Finally, comparison results are given in the fourth part, leading to propose a set of quality criteria for hyperspectral imagery. Perspectives are reported in the last section.

2. HYPERSPECTRAL IMAGERY

2.1. What is *hyperspectral* ?

A hyperspectral image is acquired by imaging spectrometers.[1] The same scene is observed at different wavelengths. The main differences between multispectral and hyperspectral imagery are in the number of bands (usually 100 to 200 bands for hyperspectral), the spectral width of these bands (narrow bands about 10 to 20 *nm*) and the fact that the bands are contiguous.

Three different manners exist to present hyperspectral images. Each of these leads to different quality criteria proposal. First of all, hyperspectral data can be viewed as 3D data, with two spatial dimensions (image) and one spectral dimension (spectrum). Hyperspectral images are therefore often represented as cubes or *hypercubes*. However, viewing hyperspectral data as cube is incomplete and neglects one important specificity of these images. This 3D view leads to consider all the three dimensions as equivalent, which is not the way they should be analyzed. Indeed, these three dimensions possess different characteristics: the hypercube is non-isotropic and there are more suitable ways to interpret hyperspectral data.

The first non-isotropic way, also the most intuitive, is to see them as a stack of images for different wavelengths. This view is coming directly from multispectral images interpretation. When considering hyperspectral data this way, typical image processing algorithms can be used. Each classical 2D image is processed, independently from the others. Results from the different wavelengths are gathered and merged.

The second way of considering hyperspectral images is to mainly focus on the spectral dimension. For every pixel, hyperspectral data can be seen as a signal or as a vector on a n_λ dimension basis (with n_λ being the number of spectral bands). When considering the data this way, applications from signal processing can used. This spectral representation leads to *typical* hyperspectral applications based on spectrum identification.

2.2. Hyperspectral applications

Applications of hyperspectral images range from agriculture (ground use) to military (detection, recognition and identification), from environmental (ocean or forestry monitoring) to geology (mineral, oil, gas exploration). The need for an important revisit capability, especially for monitoring activities is apparent. Increasing the revisit capability would be made possible with time-continuous acquisition from spaceborne instruments. Despite this wide range of applications, exploitation of the use of hyperspectral image is still at its infancy. Currently, applications use mainly the spectral information contained in hyperspectral data. Only few of them also use the spatial information. For these reasons, there is still room for improvement and the scope for new applications is wide.

This particular use of hyperspectral data, i.e. focusing more on the spectral information than on the spatial contents, creates a significant difference with the approach of traditional imagery: the spectral information (used for spectrum identification) has to be preserved. So from the perspective of near-lossless compression, spectral information has to be well preserved.

The second specific point about hyperspectral is that due to the significant amount of data, most of applications rely on computer oriented algorithms. Turning to a photo interpreter is much less common compared to traditional imagery. This particular feature makes the quality criteria developed for the HVS (Human Visual System) irrelevant.

3. QUALITY CRITERIA

3.1. Definition

In many domains, there is a need for quality criteria. For example, in classical image processing, criteria such as MSE (Mean Square Error) or SNR (Signal to Noise Ratio) are used even though it is well known that standard metrics do not reflect the perceived information loss properly. To improve the criteria for classical images, modeling for HVS has been developed.

In remote sensing field, quality criteria are used to characterize the requirements of an application from the imaging chain. The quality criteria should take into account all aspects of the data collection. Some criteria are strongly related to instrument characteristics as radiometric noise, MTF (Modulation Transfer Function), length alteration... Some others criteria, more specific, are difficult to define. This will be the case for say, a criterion representing the blocking effect of JPEG compression. In the case of hyperspectral images, since applications are particularly sensitive to the spectral dimension, suitable criteria have to closely consider it.

The purpose of this evaluation is to find the *most suitable* quality criteria for hyperspectral imagery. Such evaluations have been done for ordinary images[2] and are currently done for video sequences.[3] *Most suitable* means giving an accurate evaluation of the performances of traditional hyperspectral applications subject to a set of degradations. A *good* criterion should react to the degradations causing a performance decrease of the application; it should not react if applications are insensitive to a particular degradation.

3.2. Which criteria for hyperspectral images?

Many quality criteria have been defined in the literature. These criteria can be divided into three categories. The first one is composed of traditional criteria used in image or video processing. These criteria can be extended directly to the third dimension of hyperspectral images. In this case, the specificity of hyperspectral, explained in 2.1, is not considered. The second group of criteria is more specific to hyperspectral since they really focus on considering spectral information. In most cases, these criteria are defined on spectral vectors. Finally, the last group contains different adaptations of two advanced criteria for ordinary images. We propose to adapt them for hyperspectral images in a way such that the specificity of hyperspectral data is taken into account.

As the quality measures included in the evaluation are bivariate, they provide a measurement of a distance between I, the original hyperspectral image, and $\widetilde{I}$, the degraded one. Images are also written in a matrix form where $I(x,y,\lambda)$ denotes the value from the column x of line y in the spectral band λ. Values n_x, n_y et n_λ are the numbers of, respectively, pixels per row, samples and spectral bands. To simplify, we will denote $\sum_{x=1}^{n_x}\sum_{y=1}^{n_y}\sum_{\lambda=1}^{n_\lambda} I(x,y,\lambda)$ as $\sum_{x,y,\lambda} I(x,y,\lambda)$.

The first eight criteria, extended directly from widespread criteria, are presented below:

- Mean Square Error

$$MSE = \frac{1}{n_x n_y n_\lambda} \sum_{x,y,\lambda} \left(I(x,y,\lambda) - \widetilde{I}(x,y,\lambda) \right)^2; \tag{1}$$

- Root Mean Square Error

$$RMSE = \sqrt{MSE}; \tag{2}$$

- Relative RMSE

$$RRMSE = \sqrt{\frac{1}{n_x n_y n_\lambda} \sum_{x,y,\lambda} \left(\frac{I(x,y,\lambda) - \widetilde{I}(x,y,\lambda)}{I(x,y,\lambda)} \right)^2}; \tag{3}$$

- Maximum Absolute Difference

$$MAD = \max_{(x,y,\lambda)} \left\{ \left| I(x,y,\lambda) - \widetilde{I}(x,y,\lambda) \right| \right\}; \tag{4}$$

- Percentage MAD

$$PMAD = \max_{(x,y,\lambda)} \left\{ \frac{\left| I(x,y,\lambda) - \widetilde{I}(x,y,\lambda) \right|}{I(x,y,\lambda)} \right\} \times 100; \tag{5}$$

- Mean Absolute Error

$$MAE = \frac{1}{n_x n_y n_\lambda} \sum_{x,y,\lambda} \left| I(x,y,\lambda) - \widetilde{I}(x,y,\lambda) \right|; \tag{6}$$

- Signal to Noise Ratio

$$SNR_{(dB)} = 10.\log_{10} \frac{\sigma_I^2}{MSE}; \tag{7}$$

- Peak SNR

$$PSNR_{(dB)} = 10.\log_{10} \frac{Peak\ Signal^2}{MSE}. \tag{8}$$

The four next criteria, more specific to hyperspectral are presented below. Let μ_X be the mean of the set X and σ_X^2 be its variance. The notation $I(x,y,\cdot)$ stands for $I(x,y,\cdot) = \{I(x,y,\lambda) \mid 1 \leqslant \lambda \leqslant n_\lambda\}$. In this case $I(x,y,\cdot)$ corresponds to a vector of n_λ components.

- Maximum Spectral Similarity[4]

$$MSS = \max_{x,y} \left\{ \sqrt{RMSE\left(I(x,y,\cdot),\tilde{I}(x,y,\cdot)\right)^2 + \left(1 - corr\left(I(x,y,\cdot),\tilde{I}(x,y,\cdot)\right)^2\right)^2} \right\} \tag{9}$$

  where

$$corr\left(I(x,y,\cdot),\tilde{I}(x,y,\cdot)\right) = \frac{\frac{1}{n_\lambda - 1}\sum_{\lambda=1}^{n_\lambda}\left(I(x,y,\lambda) - \mu_{I(x,y,\cdot)}\right)\left(\tilde{I}(x,y,\lambda) - \mu_{\tilde{I}(x,y,\cdot)}\right)}{\sigma_{I(x,y,\cdot)}\sigma_{\tilde{I}(x,y,\cdot)}}; \tag{10}$$

- Maximum Spectral Angle

$$MSA = \max_{x,y} \left\{ \cos^{-1}\left(\frac{\sum_{\lambda=1}^{n_\lambda} I(x,y,\lambda).\widetilde{I}(x,y,\lambda)}{\sqrt{\sum_{\lambda=1}^{n_\lambda} I(x,y,\lambda)^2 \sum_{\lambda=1}^{n_\lambda} \widetilde{I}(x,y,\lambda)^2}} \right) \right\}; \tag{11}$$

- Maximum Spectral Information Divergence[5]

$$MSID = \max_{x,y} \left\{ \sum_\lambda (p_\lambda - \tilde{p}_\lambda) \ln\left(\frac{p_\lambda}{\tilde{p}_\lambda}\right) \right\} \tag{12}$$

  where $p_\lambda = \frac{I(x,y,\lambda)}{\|I(x,y,\cdot)\|_1}$ and $\tilde{p}_\lambda = \frac{\tilde{I}(x,y,\lambda)}{\|\tilde{I}(x,y,\cdot)\|_1}$.

- Minimum Correlation Pearson

$$Pearson = \min_{x,y} \left\{ corr\left(I(x,y,\cdot),\tilde{I}(x,y,\cdot)\right) \right\}. \tag{13}$$

The criterion developed by Wang,[6] Q, seems to give good results when applied to classical images. This criterion was also extended to video sequences. It is defined as

$$Q(X_1,X_2) = \frac{4\,\sigma_{X_1 X_2}\,\mu_{X_1}\,\mu_{X_2}}{(\sigma_{X_1}^2 + \sigma_{X_2}^2)(\mu_{X_1}^2 + \mu_{X_2}^2)}. \tag{14}$$

From this definition, three hyperspectral specific formulations are proposed in the present paper. The first adaptation is spectrum oriented, while the second one corresponds to the view of hyperspectral data as a stack of images for different wavelengths. And finally, the last adaptation tries to combine properties of both.

- Q_λ

$$Q_\lambda = \min_{(x,y)} \left\{ Q\left(I(x,y,\cdot), \widetilde{I}(x,y,\cdot) \right) \right\}; \tag{15}$$

- $Q_{(x,y)}$

$$Q_{(x,y)} = \min_{\lambda} \left\{ Q\left(I(\cdot,\cdot,\lambda), \widetilde{I}(\cdot,\cdot,\lambda) \right) \right\}; \tag{16}$$

- Q_m

$$Q_m = Q_\lambda . Q_{(x,y)}. \tag{17}$$

Then, we adapted the fidelity criterion defined by Eskicioglu[2] which gives good results when applied to gray scale images. Let the fidelity between two sets X_1 and X_2 be:

$$F(X_1, X_2) = 1 - \frac{\mathcal{L}_2^2(X_1 - X_2)}{\sigma_{X_1}^2 + \mu_{X_1}^2}. \tag{18}$$

Choosing X_1 and X_2 differently, we propose three adaptations. The first adaptation does not consider the spectral dimension separately, and this therefore corresponds to the view of hyperspectral data as an hypercube. The second one is more spectrum oriented while the last one corresponds to the stack of images at different wavelengths.

- Global Fidelity

$$F = F(I, \widetilde{I}); \tag{19}$$

- Spectral Fidelity

$$F_\lambda = \min_{(x,y)} \left\{ F\left(I(x,y,\cdot), \widetilde{I}(x,y,\cdot) \right) \right\}; \tag{20}$$

- Spatial Fidelity

$$F_{(x,y)} = \min_{\lambda} \left\{ F\left(I(\cdot,\cdot,\lambda), \widetilde{I}(\cdot,\cdot,\lambda) \right) \right\}. \tag{21}$$

All these criteria are evaluated in what follows. This list is not exhaustive but the variety of the criteria above seems efficient to orientate further researches.

4. EVALUATION METHOD

The objective of this work is to evaluate the relevance of the 18 proposed quality criteria compared to some standard applications using hyperspectral images. This problem is similar to the validation of quality criteria for classical 2D images or video sequences. In these cases, the objective is to be relevant to the human perception. In the last 20 years, many papers have tried to define a method to benchmark different criteria.[2,7] An adaptation from the *Video Quality Expert Group* (VQEG)[3] will be used here.

In the case of hyperspectral images, human experts seldom visualize hyperspectral data. Most of the time, data are directly processed by algorithms. For this reason, the *subjective* evaluation is done through different applications. However, unless we can expect all human observers to react in a similar way when subject to the same degradation, such expectations may not be true for different hyperspectral applications. Each application has a different sensitivity to different degradations.

First, the images used during the evaluation process are presented and their main characteristics are highlighted. Then, the reference applications used to benchmark quality criteria are detailed. Thereafter, degradations are analyzed and interpreted. And finally, the process of evaluation is defined.

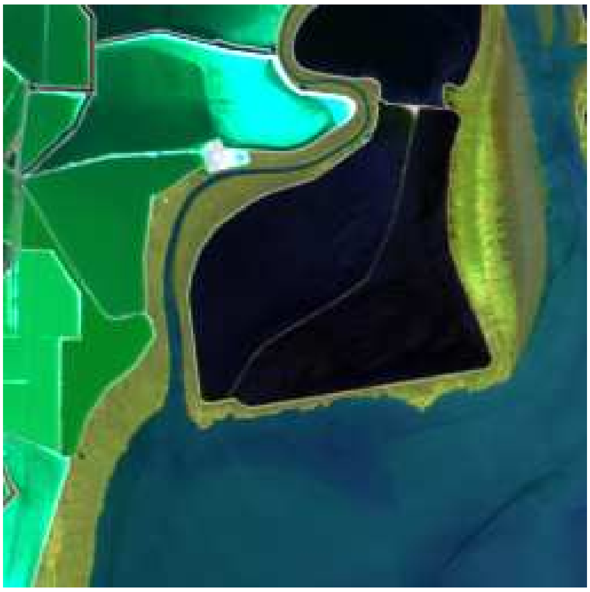

Figure 1. Moffett2 with uniform zones

Figure 2. Moffett3 uneven zones

4.1. Images

The SNR characteristics of the images chosen for the simulation is important. With a low SNR, the added noise would be hidden under the instrument noise. Since the purpose of this study is to propose quality criteria for near-lossless compression, the impact of a very light noise has to be evaluated. Among all available hyperspectral images, Aviris images from NASA/JPL present good SNR characteristics (about 600:1) and therefore will be used in the following simulations.

To make the results more accurate, the simulations are done on two parts of the Moffett Field site (California) with different properties. The part denoted *moffett2* (Fig. 1) contains large uniform zones (salt evaporator and sea) whereas *moffett3* (Fig. 2) is more uneven (roads and buildings).

4.2. Reference applications

Many applications for hyperspectral are based on spectrum matching. This is the case of supervised classification. In this case, pixels are classified according to the distance between their spectrum and a spectrum of reference. The reference spectrum can come either from a spectral library containing samples of spectrum for different materials, or from a region of interest defined on the image to classify. The measured distance can be defined in different ways. In the present paper, a classification based on user-defined region of interest to compute the statistical properties for each class will be used. Three different classification processes will be studied, namely, *Spectral Angle Mapper* (SAM), which uses the spectral angle as a distance measure, *Mahalanobis classification*, using the Mahalanobis distance and a *Maximum Likelihood Classification*.[8] For each classification, statistics on the regions of interest are calculated: mean spectrum, variance, covariance matrix... Pixels are classified based on the properties of their spectrum. In the case of the *Spectral Angle Mapper*, a threshold is defined. Pixels which are too far from the classes will stay unclassified.

Without ground reference to benchmark the performances of the classification (which is not the goal here), only classifications variations will be estimated. The score given by the classification will be the amount of misclassified pixels compared to the reference classification for the original image.

4.3. Degradation simulations

To provide accurate results, different degradations are applied to the hyperspectral images. Four different degradations at different levels are used. These four types of degradation represent the typical degradations which occurs: Gaussian white additive noise, spatial or spectral smoothing, Gibbs effect and misregistration.

The first type of degradation applied to hyperspectral images is an additive Gaussian white noise with different variances. This noise models the instrumental photonic or electronic noise. A random noise of a given variance is added to the entire image.

A smoothing filter can be applied to the image, either on the spatial dimensions, the spectral dimension or both. During the acquisition process smoothing can come from an MTF default (size of the impulse response), from a particular type of compression (wavelets for example tend to blur the image), or from a lower resolution. To apply this degradation, a low-pass filter is used. The slope of the filter can be adjusted to reduce the effect.

The third degradation is a modeling of the Gibbs effect causing ringing around sharp changes. This effect can appear during the post-treatments when applying low-pass filters. This effect is created using a modified Wiener filter only for the spatial direction.

The last applied degradation is the misregistration. The misregistration denotes a bad alignment between two different bands. The two different image planes are not properly superimposable. Even if the misregistration is limited for hyperspectral images compared to multispectral images, it is interesting to see its influence on quality criteria. This degradation is done shifting randomly the spectral plane in x and y directions. Random values of shifting are casted for each spectral plane. The shift is between 0 and 1 pixel on both directions and can be limited. The calculation of the new values is done by cubic interpolation.

4.4. Evaluation process

Extensive degradation situations have been simulated. For each *situation* (namely one degradation of a certain level applied to one image), all quality criteria are calculated and all classification results are measured. Usually, to compare these results, a correlation is applied. Brill[9] developed a more complete method for the evaluation of quality criteria within the frame of video sequences. However, his method cannot be directly applied for the hyperspectral case due to the difference between human observers and the classification algorithms. Whereas among human observers, variability is present but small, classification algorithms are deterministic but can produce very different results depending on which algorithm is applied.

Thereafter, Brill's method is modified. For each situation i corresponding to one degradation of a certain level applied to one image, a score is computed for each quality criterion (O_i, objective score) and for each application (S_i, *subjective* score). For example, the image *moffett2* with a white noise of variance 60 is one situation. For this situation, every quality criterion is computed and every application is processed. The curve representing the application performance versus the quality criteria (point (O_i, S_i)) enables us to spot the more sensitive degradations for a given quality criterion.

5. SIMULATION RESULTS

5.1. Curves

For each curve (O_i, S_i), the abscissa represents the values of the quality criteria while the ordinate represents the value for the application. To make the curves easier to read, the value of the quality criteria representing the maximum possible quality is on the left side. Likewise, the classification value corresponding to the best quality (no classification error) is on the lower side. Then, the origin of the graph is on the lower left corner. Different symbols are used to plot the different degradations (Tab. 1).

A quality criterion is considered reliable if the dispersion of the curve is low and in an oblique direction as in Fig. 4. This case means that when the amount of error increases for the classification, the quality criterion reacts to it in about the same proportion for every degradation. On the other hand, if the points are scattered, the criterion neglects some types

White Noise	$+$
Spectral Smoothing	$*$
Spatial Smoothing	$\times$
Mixed smoothing (spatial and spectral)	$\square$
Gibbs effect	$\Diamond$
Misregistration	$\triangle$

Table 1. Symbols used to represent the different degradations

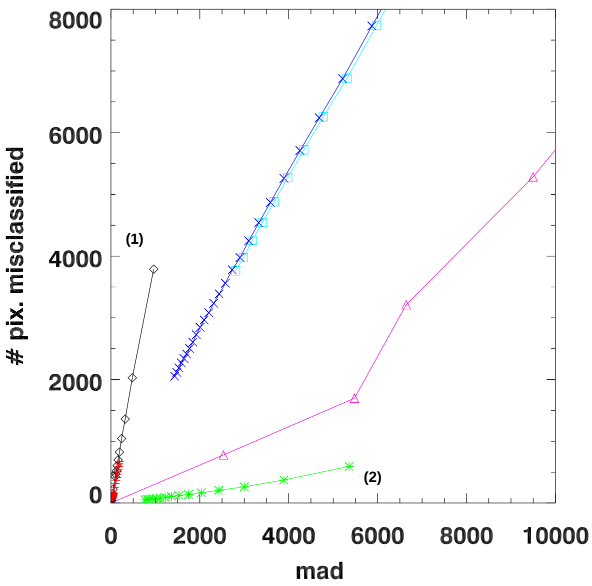

Figure 3. MAD (Eq. 4) vs. SAM classification: underestimation of the Gibbs effect (1) and overestimation of the spectral smoothing (2). This is also the criterion the most sensitive to misregistration.

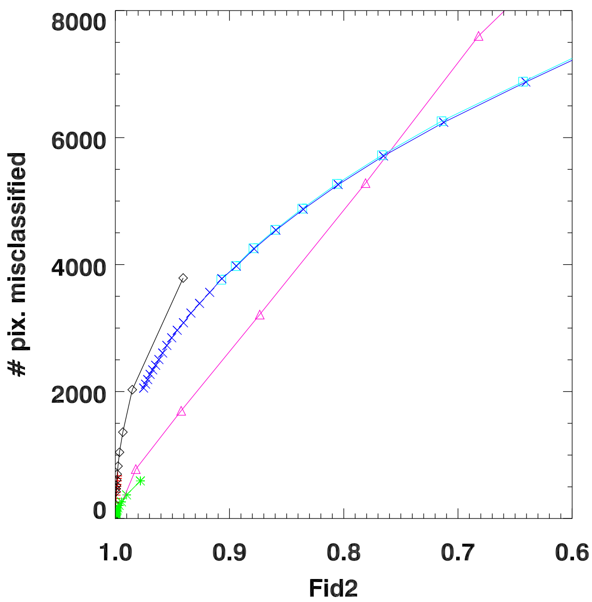

Figure 4. F_λ (Eq. 20) vs. SAM classification: dispersion for all degradation is low, the reaction of this criterion is similar with the classification. This criterion gives a reliable estimation of the degradation.

of degradations and overestimates some others. When the criterion response is concentrated around a vertical direction (Fig. 3 curve 1), the criterion does not react for a degradation which affects the performances. On the other hand, when the response is concentrated around a horizontal direction (Fig. 3 curve 2), the criterion overestimates the impact of the degradation on the application performances (which are almost not affected).

As we can see in Fig. 4, the F_λ criterion seems to give a good estimation of the impact of all degradations on SAM classification. However, as different classification methods have different properties, it is not possible to keep only one criterion for an accurate estimation.

Actually two types of criteria can be useful. One type could be criteria which react directly like the application, such as F_λ. And the other type could be criteria that neglect some degradations and overestimate others. With a panel of well-chosen criteria, it even should be possible to define the nature of the degradation.

While SAM classification is not sensitive to the presence of white noise in the image, Mahalanobis classification and Maximum Likelihood classification are very sensitive to additive white noise. Another difference appears in the case of spectral smoothing degradation with a threshold effect: for light spectral smoothing, classification results are not altered, but above a certain level, the effect is important.

All the curves for all the criteria in all the tested situations can be obtained at http://www.enseeiht.fr/∼christophe/quality. The obtained results vary greatly from one criterion to the other. Some overestimate or underestimate the impact of a set of degradations. The spectral smoothing, for example, is overestimated by most criteria. Except F_λ and Q_λ, all criteria (Fig. 5 for example) overestimate its impact. In the case of Mahalanobis classification, the spectral smoothing causes a threshold effect (Fig. 6). Some criteria such as PMAD, RRMSE, MAE and MSID are very sensitive to a white noise presence whereas SAM classification does not react to this (Fig. 7). $Q_{(x,y)}$ is completely insensitive to the degradation affecting spacial planes like Gibbs effect or spatial smoothing, even if these degradations have an impact on the spectrum (Fig. 8).

5.2. Which quality criteria?

From the evaluation curves, some properties of criteria can be extracted. They are presented in Tab. 2. Five criteria, namely, RRMSE, MAE, MAD, $Q_{(x,y)}$, F_λ, can be computed on each hyperspectral image to give an accurate estimation of the nature of the degradation and of their intensity. Moreover, the F_λ alone gives a good estimation of the influence of the degradation

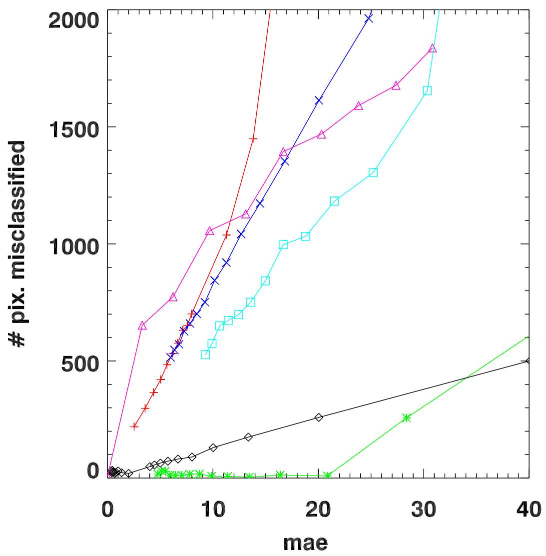

Figure 5. MAE (Eq. 6) vs. Mahalanobis classification: MAE is the most sensitive criterion to the presence of Gibbs effect

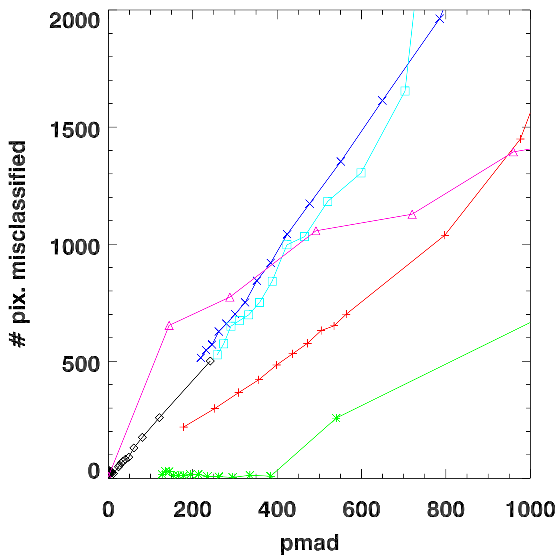

Figure 6. PMAD (Eq. 5) vs. Mahalanobis classification: apparition of a threshold effect for the spectral smoothing

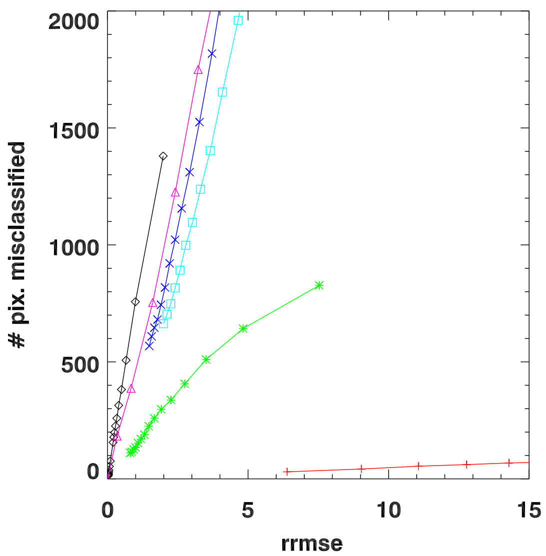

Figure 7. RRMSE (Eq. 3) vs. SAM classification: RRMSE is overreacting to the presence of white noise

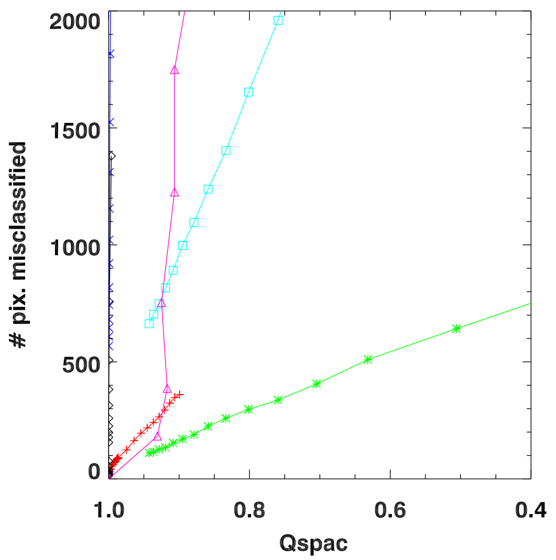

Figure 8. $Q_{(x,y)}$ (Eq. 16) vs. SAM classification: no sensitivity to the presence of spatial smoothing

when the application is SAM classification. These results have been intensively tested and are stable when applied to two images with different properties.

	White Noise	Spectral Smooth.	Spatial Smooth.	Gibbs effect	Misregist.
+ sensitive	RRMSE	$Q_{(x,y)}$	F_λ	MAE	MAD
- sensitive	F_λ	F_λ	$Q_{(x,y)}$	$Q_{(x,y)}$	RRMSE

Table 2. Most and least sensitive quality criterion for each degradation

These results are obtained observing the curves. The most sensitive and the least sensitive criteria for each degradation are highlighted. It would be usefull to repeat the benchmark on other applications, even if the chosen classifications are representative of the typical applications. The authors will be glad to repeat the simulations on other existing applications.

6. PERSPECTIVES

The purpose of this study was originally to find suitable quality criteria to estimate accurately the influence of compression noise on hyperspectral applications. However, the possible use of the defined criteria goes beyond the scope of the compression. As we seen before, the five defined criteria can be used to characterize the nature of the noise affecting one image. For example, when combining the RRMSE and the $Q_{(x,y)}$, if the RRMSE indicates a strong degradation whereas the $Q_{(x,y)}$ does not react, it can be concluded that the degradation is probably similar to a white noise. Combining the five defined criteria, more precise estimation can be obtained.

The studied criteria are all bivariate, which means that they require the original image to be calculated. When the problem is to evaluate the quality of the complete imagery chain, the original image going through the instrument is not obtainable. This particular problem can be overcome using image modeling. We can use the instrument to realize an acquisition on a test site, while, knowing the ground truth, the image which should be received can be modeled. Finally comparison can be done using the quality criteria in order to determine the major point to improve in the instrument conception.

ACKNOWLEDGMENTS

This work has been carried out under the financial support of *Centre National d'Etudes Spatiales* (CNES), *Office National d'Etudes et de Recherches Aérospatiales* (ONERA) and *Alcatel Space*. The authors wish to thank *NASA/JPL* and *European Space Agency* (ESA) for providing the hyperspectral images used during the experiments.

REFERENCES

1. D. Landgrebe, "Hyperspectral image data analysis," *IEEE Signal Processing Magazine* **19**, pp. 17–28, Jan. 2002.
2. A. M. Eskicioglu and P. S. Fisher, "Image quality measures and their performance," *IEEE Transactions on communications* **43**, pp. 2959–2965, Dec. 1995.
3. VQEG, "Final report from the video quality experts group on the validation of objective models of video quality assessment, phase II," tech. rep., VQEG, Aug. 2003. http://www.vqeg.org.
4. S. Rupert, M. Sharp, J. Sweet, and E. Cincotta, "Noise constrained hyperspectral data compression," in *Geoscience and Remote Sensing Symposium, 2001*, **1**, pp. 94–96, IEEE, July 2001.
5. B. Aiazzi, L. Alparone, S. Baronti, C. Lastri, L. Santurri, and M. Selva, "Spectral distortion evaluation in lossy compression of hyperspectral imagery," in *Geoscience and Remote Sensing Symposium 2003*, **3**, pp. 1817–1819, IEEE, Jul. 2003.
6. Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Processing Letters* **9**, pp. 81–84, March 2002.
7. N. Damera-Venkata, T. Kite, W. Geisler, B. Evans, and A. Bovik, "Image quality assessment based on a degradation model," *IEEE Transactions on Image Processing* **9**, pp. 636–650, Apr. 2000.
8. J. A. Richards, *Remote Sensing Digital Image Analysis*, Springer-Verlag, Berlin, 1999.
9. M. H. Brill, J. Lubin, P. Costa, S. Wolf, and J. Pearson, "Accuracy and cross-calibration of video quality metrics: new methods from ATIS/T1A1," *Signal Processing: Image Communication* **19**, pp. 101–107, 2004.