Full Reference Image Quality Assessment based on Local Statistics[1]

Santiago Aja-Fernández
LPI, ETSI Telecomunicación, Universidad de Valladolid, Spain
Email: sanaja@tel.uva.es
Raúl San José Estépar
LMI, Brigham and Women's Hospital, Harvard Medical School.
Carlos Alberola-López
LPI, ETSI Telecomunicación, Universidad de Valladolid, Spain

Abstract

Assessment of image quality based on structural information is a relevant topic in the image processing field. A new method to assess image quality is presented. It is based on the comparison of first and second order moments of the local variance distribution of two images. This new quality index is better suited to assess the non-stationarity of images, therefore it explicitly focuses on the image structure. A second index is also proposed to be used for low noise cases. This new quality indexes are compared to state-of-the art objective methods for a wide range of signal transformations and images.

Key words: Quality assessment, local variance

INTRODUCTION

Images can suffer distortion due to several sources, from the acquisition process itself to compression, transmission through noisy channels and others. On the other hand, images can also undergo quality improvement processes, like enhancement or restoration techniques (1). In each case it is useful to quantify the *quality* of the resulting image. One easy way to do it is by using a reference image to carry out this task. These approaches are known as *full-reference methods* (2). The most straightforward parameters are those based on pixelwise error measurements, like MSE (3), (4) and other related parameters. Methods that do not use a reference image are known as *blind quality assessment* methods (5), (6), (7), (8). Aswell, a small set of parameters of the reference image may be used for quality assessment, using a *reduced-reference* criterion (2). In this paper we will focus on the first kind.

The sources of degradation that an image may suffer are wide and diverse. An ideal quality assessment method should be able to cope and to quantify any kind of distortion. However this may be quite a hard task —not to say impossible— and probably application dependent, since the importance of a specific type of degradation is different depending on the purpose of the image, i.e., a particular noise level may be acceptable in home pictures but may lead to poor results in a segmentation application; some blur of the edges may lead to a critical information lost in MRI whereas the same process may even be able to ease the interpretation of an ultrasound image.

Recent methods for quality assessment rely on the analysis of the structural information in the image. These methods have proved of great interest for very different kinds of images, ranging from natural scenes to medical scenarios. However, due to the great variety of possible degradations one may think of situations in which the information provided by a specific measure does not match a subjective quality judgement. Methods may have a bias towards the *image statistic* of the structural measure on which the design is grounded.

In this paper, we present a new method based on the structural information of the image and, specifically, on the statistics of the sample local variance; it is intended to penalize degradations in which the structural content is filtered out by accounting for the non-stationary content in the image. Additionally, it intends not to over-penalize noise as long as noise content does not obscure structure. This new method can be seen as a new stand-alone index, or as a complement to other existing methods. This paper is an extension of our preliminary proposal presented in (9).

[1]TECH-LPI2014-01. This work is an original Technical Report of the LPI, Universidad de Valladolid, Spain, Jan. 2014. www.lpi.tel.uva.es/~santi. Citation:
Santiago Aja-Fernández, Raúl San José Estépar, Carlos Alberola-López, Full Reference Image Quality Assessment based on Local Statistics, *Tech Report of the LPI*, TECH-LPI2014-01, Universidad de Valladolid, Spain, Jan. 2014. www.lpi.tel.uva.es/~santi

FULL-REFERENCE QUALITY ASSESSMENT

Full-reference methods for quality assessment are those in which a signal is compared to a *ground truth* image, i.e. a *golden standard*. Within these methods, the most frequently used are those error based methods, as the Mean Squared Error (MSE) (3), (4). Let $I(\mathbf{x})$ be the *ground truth image* and $J(\mathbf{x})$ an image we want to compare with the former. The MSE is defined

$$\mathrm{MSE}(I,J) = \frac{1}{|\Omega|}\sum_{\mathbf{x}\in\Omega}\left(I(\mathbf{x})-J(\mathbf{x})\right)^2 \qquad [1]$$

being $|\Omega|$ the cardinal of space Ω, i.e., the number of elements in image $I(\mathbf{x})$. Although the MSE gives a measure of pixelwise similarity between the images, it does not explicitly take into account any structural information in the images or any sort of subjective measure. Some variations are sometimes considered (4), such as the Structural Content (SC)

$$\mathrm{SC}(I,J) = \frac{\frac{1}{|\Omega|}\sum_{\mathbf{x}\in\Omega}[I(\mathbf{x})]^2}{\frac{1}{|\Omega|}\sum_{\mathbf{x}\in\Omega}[J(\mathbf{x})]^2} \qquad [2]$$

or the Peak Signal to Noise Ratio (PSNR) in dB units

$$\mathrm{PSNR}(I,J) = 10\log_{10}\left(\frac{\left(\max_{\mathbf{x}} I(\mathbf{x})\right)^2}{\frac{1}{|\Omega|}\sum_{\mathbf{x}\in\Omega}\left(I(\mathbf{x})-J(\mathbf{x})\right)^2}\right). \qquad [3]$$

The limitations of such methods have been widely reported in the literature (see (10) for example). Consequently, some additional variations of the MSE have also been used in order to better deal with the features of the Human Visual System (3), (4), like the Laplacian MSE

$$\mathrm{LMSE}(I,J) = \frac{\frac{1}{|\Omega|}\sum_{\mathbf{x}\in\Omega}\left(H(I(\mathbf{x}))-H(J(\mathbf{x}))\right)^2}{\frac{1}{|\Omega|}\sum_{\mathbf{x}\in\Omega}\left(H(I(\mathbf{x}))\right)^2} \qquad [4]$$

where the images are previously high-pass filtered with the high-pass operator $H(\cdot)$. In (11), Przelaskowski proposes a vector of six components, mainly based on errors between the ground truth and the degraded [compressed] image, such as the average pixel error, correlated errors or preservation of high contrast edges. Methods proposed in (12) are also mainly based on error measurements. Other approaches, as the one in (13), assess the quality using a degradation model. In (14) a new index is proposed, namely, the objective Picture Quality Scale (PQS), basically intended to measure the degradation in coding and compression of images. It takes into account properties of visual perception of both global features and of disturbances. It turns out to be bounded, being the maximum value 5.797, obtained when an image is compared with itself. Experiments show that although it is a good measure when dealing with compression, it is not so good a measure for other sources of degradation (15). Recently, some methods based on *Natural Scene Statistics* have been reported (16), (17).

In (2) Wang *et al.* proposed a full-reference quality assessment method based on the structural similarity of two images, the so-called Structural Similarity (SSIM) index. The method is a modification of their Quality Index, originally proposed in (18). As of today, this method has proved to be versatile and robust in many different environments (15). It uses three levels of [pixelwise] comparison:

1) Luminance comparison:

$$l(I,J) = \frac{2\mu_I(\mathbf{x})\mu_J(\mathbf{x})+C_1}{\mu_I^2(\mathbf{x})+\mu_J^2(\mathbf{x})+C_1}$$

with $\mu_I(\mathbf{x})$ and $\mu_J(\mathbf{x})$ the local means of the images $I(\mathbf{x})$ and $J(\mathbf{x})$ respectively, and C_1 a constant.

2) Contrast comparison:

$$c(I,J) = \frac{2\sigma_I(\mathbf{x})\sigma_J(\mathbf{x})+C_2}{\sigma_I^2(\mathbf{x})+\sigma_J^2(\mathbf{x})+C_2}$$

with $\sigma_I(\mathbf{x})$ and $\sigma_J(\mathbf{x})$ the local standard deviations of the images $I(\mathbf{x})$ and $J(\mathbf{x})$, respectively, and C_2 a constant.

3) Structure comparison:

$$s(I,J) = \frac{\sigma_{IJ}(\mathbf{x}) + C_3}{\sigma_I(\mathbf{x})\sigma_J(\mathbf{x}) + C_3}$$

with $\sigma_{IJ}(\mathbf{x})$ the local covariance between the images $I(\mathbf{x})$ and $J(\mathbf{x})$, and C_3 a constant.

The local SSIM index then is defined as

$$\text{SSIM}(I,J) = [l(I,J)]^{\alpha} \cdot [c(I,J)]^{\beta} \cdot [s(I,J)]^{\gamma}\,. \quad [5]$$

with α, β and γ weights in the interval $[0,1]$. The overall value is obtained using the mean of the local SSIM (with acronym MSSIM):

$$\text{MSSIM}(I,J) = \frac{1}{|\Omega|}\sum_{\mathbf{x}\in\Omega}\text{SSIM}(I(\mathbf{x}),J(\mathbf{x})) \quad [6]$$

Some variations of the original methods have been proposed elsewhere, like using a weighted sum instead of the mean (19).

$$\text{WSSIM}(I,J) = \frac{\sum_{\mathbf{x}\in\Omega} W(\mathbf{x})\text{SSIM}(I(\mathbf{x}),J(\mathbf{x}))}{\sum_{\mathbf{x}\in\Omega} W(\mathbf{x})} \quad [7]$$

with $W(\mathbf{x})$ a weighting function that can be, for example:

$$W(\mathbf{x}) = \sigma_I^2(\mathbf{x}) + \sigma_J^2(\mathbf{x}) + C_2$$

A somehow different approach is the one by Weken *et. al*; they use fuzzy similarity measures as a way to compare two images. In (20), (21), (22) many distances have been defined, tested and used, either over the images themselves or over their histograms. For instance, using the fuzzy Minkowski distance

$$S_1(I,J) = 1 - \left(\frac{1}{|\Omega|}\sum_{\mathbf{x}\in\Omega}|I(\mathbf{x}) - J(\mathbf{x})|^r\right)^{1/r} \quad [8]$$

or a modified version

$$M_3(I,J) = 1 - \frac{\sum_{\mathbf{x}\in\Omega}|I(\mathbf{x}) - J(\mathbf{x})|}{\sum_{\mathbf{x}\in\Omega}(I(\mathbf{x}) + J(\mathbf{x}))} \quad [9]$$

To be considered as fuzzy sets, the images must be normalized by their maximum value. The measures are bounded, giving an index in the interval $[0,1]$.

Some of the methods for image quality assessment presented in this section have been recently statistically evaluated in (15). The quality given by the algorithms has been compared to the subjective observer judgements. An interesting conclusion of this study is that the field still needs new methods to evaluate image quality in order to reduce the gap between human-based and machine-based assessment. The results of this study also seem to favor algorithms based on structural information, like SSIM, as well as natural scene statistics, over other measurements.

IMAGE QUALITY ASSESSMENT BASED ON LOCAL VARIANCE

In (23) some properties are considered for quality assessment methods based on similarity measures. Let $I(\mathbf{x})$ be the *ground truth*, $J(\mathbf{x})$ the image we want to compare with the former and $Q()$ a quality comparison method. $Q(I,J)$ must have the following properties:

1) Reflexivity: $Q(I,I) \geq Q(I,J)$[1]. If we assume a bounded method assessment method with results within $[0,1]$, then $Q(I,I) = 1$.

[1]This property applies only for distorted images, since enhanced ones should yield the opposite result. However, its application in this case is straightforward by just considering the distortion with the enhanced image as reference.

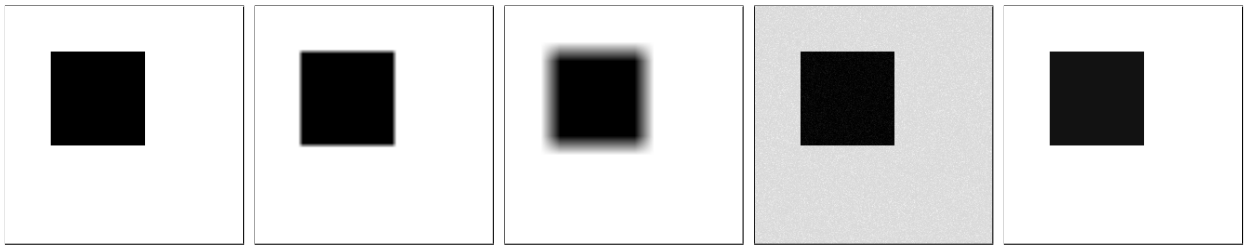
Distorsion	MSE	PSNR	MSSIM	PQS	M3
Blur 5×5	160.04	26.09	**0.96**	0	0
Blur 21×21	692.49	19.73	0.87	0	0
White Noise	943.09	**18.39**	0.63	0	0
Constant	100	28.13	0.86	0	0

TABLE I. Quality assessment using different measures for the black square experiment in Fig. 1.

2) Symmetry: $Q(I, J) = Q(J, I)$.
3) Reaction to artifacts such as noise, blurring, blocking, etc.: a good similarity measure should not overreact to artifacts, and should be decreasing with respect to an increasing distortion level.
4) Reaction to enlighting and darkening: The output of a enlighted or darkened image must be high, as it is similar to the original one. If they are applied so that the image is too affected to see the details, one also expects a decreasing behavior with respect to an increasing enlighting or darkening percentage.

Bearing these properties in mind, in the following sections we will describe a new full-reference method to assess the quality of an image.

A previous experiment



(a) (b) (c) (d) (e)

Figure 1. Synthetic experiment: Black square (256 gray levels). (a) Original Image, (b) Blurred image using a square 5×5 window, (c) Blurred Image using a square 21×21 window, (d) Image with additive Gaussian noise with 0 mean and $\sigma = 5$, (e) Image plus constant 10.

From the methods reviewed in the previous section, only those based on fuzzy measures, the PQS and the MSSIM are bounded. Although these indexes have been shown to be useful when measuring many different distortions in images, cases may arise in which the quality measure obtained does not properly match a subjective judgment based on the visual information. As an example, consider the synthetic image in Fig. 1-(a), degraded from different sources:

- The image is blurred via convolution with a 5×5 averaging kernel, Fig. 1-(b).
- The image is blurred via convolution with a 21×21 averaging kernel, Fig. 1-(c).
- The image is corrupted by Gaussian noise with 0 mean and $\sigma = 5$, Fig. 1-(d).
- Finally, a constant (10) is added to the image, Fig. 1-(e).

The quality of the degraded images is assessed using different methods: MSE, PSNR, MSSIM, PQS (normalized by 5.797) and the fuzzy measure in eq. ([9]) over the histogram. Results are on Table I.

From this particular example we can deduce that the PQS and the M3 due to a very fast decay of the functions on upon which they are defined claim that the images are totally different, when reality is far from this. Error-based methods, MSE and PSNR, give measures that are not totally consistent. The MSSIM seems to be the only one that gives a bounded measure. However, this index considers some sources of degradation more important than others, i.e., there exists a bias towards some features of the image. For instance, blur is taken as a minimal degradation, although for many applications it may constitute an important structural loss; on the other hand, white noise is seen as a substantial degrading effect although, as a matter of fact, noisy structures may be *clearer* to the human eye than the blurred ones (when, for instance, identifying organs in ultrasound images). If a quality index is to be used to test restoration filters or enhancement methods, both blur and robustness to noise must be taken accounted for.

Some other related examples will be shown in section IV. In order to reduce this bias indicated above, alternative quality measures should be conceived, for instance, that rely on different sources of structural information. In following sections new methods based on local statistics –mainly local variance– are introduced.

Image degradation and local variance distribution

The distribution of the local variance in an image has been previously used as a way to study the image structure (24) or to estimate the level of noise (25), (26), (27). It keeps important information of the structural content in the image, and any degradation will affect such distribution. In this section we will show how it is affected by different kind of distortions.

The local variance of an image $I(\mathbf{x})$ is defined as

$$\sigma_I^2(\mathbf{x}) = E\left\{I^2(\mathbf{x})\right\} - E\left\{I(\mathbf{x})\right\}^2 \quad [10]$$

being $E\{I(\mathbf{x})\}$ the local mean of the image. It may be estimated by the *sample local variance* using a square neighborhood $\eta(\mathbf{x})$ centered about the pixel under analysis as

$$\hat{\sigma}_I^2(\mathbf{x}) = \frac{1}{|\eta(\mathbf{x}) - 1|} \sum_{\mathbf{p}\in\eta(\mathbf{x})} \left(I(\mathbf{p}) - \hat{\mu}_I(\mathbf{p})\right)^2 \quad [11]$$

with $\mu_I(\mathbf{x})$ the sample local mean of image I defined as

$$\hat{\mu}_I(\mathbf{x}) = \frac{1}{|\eta(\mathbf{x})|} \sum_{\mathbf{p}\in\eta(\mathbf{x})} I(\mathbf{p})$$

If the estimation is to be done using a weighing function $\omega_{\mathbf{p}}$ (such as a Gaussian function) in neighborhood $\eta(\mathbf{x})$ centered about the pixel under analysis, then

$$\hat{\sigma}_I^2(\mathbf{x}) = \frac{\sum\limits_{\mathbf{p}\in\eta(\mathbf{x})} \omega_{\mathbf{p}} \left(I(\mathbf{p}) - \hat{\mu}_I(\mathbf{p})\right)^2}{\sum\limits_{\mathbf{p}\in\eta(\mathbf{x})} \omega_{\mathbf{p}}} \quad [12]$$

with

$$\hat{\mu}_I(\mathbf{x}) = \frac{\sum\limits_{\mathbf{p}\in\eta(\mathbf{p})} \omega_{\mathbf{p}} I(\mathbf{p})}{\sum\limits_{\mathbf{p}\in\eta(\mathbf{p})} \omega_{\mathbf{p}}}. \quad [13]$$

The size of the neighborhood $\eta(\mathbf{x})$ is related to the scale of the image structures expected for a particular application.

In this paper we assume that all the local variances in the image are identically distributed random variables. Let V_I denote any of the variables $\sigma_I^2(\mathbf{x})$. Then, the distribution of the local Variance $p_{V_I}(v)$ may be approximated by the histogram of the sample local variance $\sigma_I(\mathbf{x})$. For the sake of illustration, a picture from a natural scene –see Fig. 2– will be degraded using different common sources:

1) White Gaussian noise with zero mean and $\sigma = 20$ is added to the image. This is the kind of thermal noise that may be found in many applications.
2) The image is corrupted with salt and pepper (S&P) noise with 5% density. This kind of noise may be caused by errors in the data transmission, malfunctioning pixels or dust in sensors.
3) The image is corrupted with Rician noise with $\sigma_n = 20$ (27). The Rician distribution may be found in noisy magnetic resonance data, and it is also a model for speckle.
4) The image is blurred using a 11×11 Gaussian kernel with $\sigma = 10$.
5) The image is averaged using a 21×21 square window.
6) To analyze the effect of compression, the image is compressed using a commercial JPEG algorithm with high compression rate.

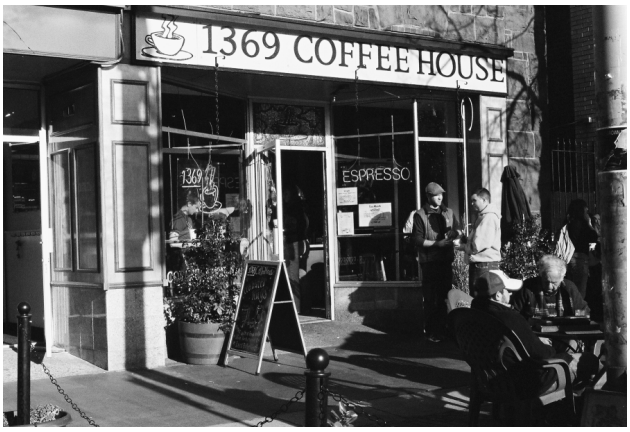
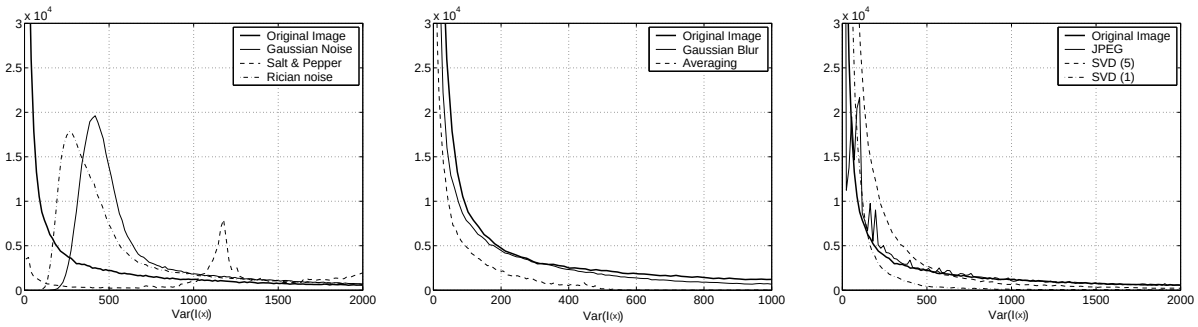


Figure 2. Image used for illustration of behavior of the local variance distribution.



(a) Noise (b) Blur (c) Compression

Figure 3. Effects of different sources of degradation over the local variance distribution of an image. (a) Effect of noise: Additive Gaussian noise ($\sigma_n = 20$), salt and pepper (5%) and Rician ($\sigma_n = 20$). (b) Effect of blur: Gaussian blur ($\sigma = 10$ and 11×11 kernel) and averaging (using a 21×21 window). (c) Effect of compression, JPEG (with high compression rate) and Singular Value Decomposition (5 most significant eigenimages and most significant eigenimage)

7) Finally, to analyze a spectral decomposition, the Singular Value Decomposition (SVD) (28) of the image is done, and the most and 5 most significant eigenimages are alternatively selected.

For the original image as well as for each of the degradations, the distribution of the local variance is estimated and depicted in Fig. 3. As pointed out in (27), a natural scene image shows a typical exponentially-decreasing histogram shape, with its maximum close to the origin. When the image is corrupted with additive or Rician noise, the maximum of the histogram shifts rightwards, changing the mean, the median and the mode of the distribution (26), (27). The effect of adding S&P noise is a total change in the shape of the distribution. On the other hand, blurring and averaging the image has the effect of decreasing the variance of the distribution. Finally, a JPEG compression and the SVD also change the shape and variance of the distribution. Note that the JPEG blocking effect translates in some *peaks* along the histogram.

Quality Index based on Local Variance

The proposed new index is based on the assumption that a great amount of the structural information of an image is encoded in the distribution of its local variance. Although the local variance itself has been taken accounted for in other indexes, like the SSIM, its statistics, however, have been widely ignored. In our opinion this information lost is particularly relevant, as the example in the previous section has just pointed out. Specifically, we will make use of the

mean of of V_I

$$\mu_{V_I} = E\left\{\sigma_I^2(\mathbf{x})\right\}$$

estimated as

$$\hat{\mu}_{V_I} = \frac{1}{|\Omega|}\sum_{\mathbf{x}\in\Omega}\hat{\sigma}_I^2(\mathbf{x}) \quad [14]$$

The standard deviation of V_I, defined as

$$\sigma_{V_I} = \left(E\left\{(\sigma_I^2(\mathbf{x}) - \mu_{V_I})^2\right\}\right)^{1/2}$$

and estimated as

$$\hat{\sigma}_{V_I} = \left(\frac{1}{|\Omega|-1}\sum_{\mathbf{x}\in\Omega}\left(\hat{\sigma}_I^2(\mathbf{x}) - \hat{\mu}_{V_I}\right)^2\right)^{1/2} \quad [15]$$

will also be considered. Additionally, cross information between the two images will enter the assessment measurement as well by using the covariance between variables V_I and V_J, i.e.,

$$\sigma_{V_IV_J} = E\left\{(\sigma_I^2(\mathbf{x}) - \mu_{V_I})(\sigma_J^2(\mathbf{x}) - \mu_{V_J})\right\} \quad [16]$$

which will be estimated by means of

$$\hat{\sigma}_{V_IV_J} = \frac{1}{|\Omega|-1}\sum_{\mathbf{x}\in\Omega}\left(\hat{\sigma}_I^2(\mathbf{x}) - \hat{\mu}_{V_I}\right)\left(\hat{\sigma}_J^2(\mathbf{x}) - \hat{\mu}_{V_J}\right) \quad [17]$$

We define the *Quality Index based on Local Variance* (QILV) between two images I and J as

$$\text{QILV}(I,J) = \frac{2\mu_{V_I}\mu_{V_J}}{\mu_{V_I}^2 + \mu_{V_J}^2}\cdot\frac{2\sigma_{V_I}\sigma_{V_J}}{\sigma_{V_I}^2 + \sigma_{V_J}^2}\cdot\frac{\sigma_{V_IV_J}}{\sigma_{V_I}\sigma_{V_J}} \quad [18]$$

Note that although there is a purposedly great similarity between eq. ([18]) and the SSIM index, the latter relies on the mean of the local statistics of the images, the former deals with the [global] statistics of the local variances of the images.

The first term in eq. ([18]) carries out a comparison between the means of the local variances of both images. The second one compares the standard deviation of the local variances. This term is related with the *blur* and the *sharpness* of the image. The third term is the one to introduce cross information in the two images. To avoid computational problems with small values, some constants may be added to every term in eq. ([18])

$$\text{QILV}(I,J) = \frac{2\mu_{V_I}\mu_{V_J} + C_4}{\mu_{V_I}^2 + \mu_{V_J}^2 + C_4}\cdot\frac{2\sigma_{V_I}\sigma_{V_J} + C_5}{\sigma_{V_I}^2 + \sigma_{V_J}^2 + C_5}\cdot\frac{\sigma_{V_IV_J} + C_6}{\sigma_{V_I}\sigma_{V_J} + C_6} \quad [19]$$

In order to make the index more sensible to certain kind of degradations, each of the three components may we weighed by a different positive exponent α, β and γ:

$$\text{QILV}(I,J) = \left[\frac{2\mu_{V_I}\mu_{V_J}}{\mu_{V_I}^2 + \mu_{V_J}^2}\right]^\alpha\cdot\left[\frac{2\sigma_{V_I}\sigma_{V_J}}{\sigma_{V_I}^2 + \sigma_{V_J}^2}\right]^\beta\cdot\left[\frac{\sigma_{V_IV_J}}{\sigma_{V_I}\sigma_{V_J}}\right]^\gamma \quad [20]$$

We can now compare the previous results for the black square in Fig.1 and Table I with this new index in Table II. For easier reading, the MSSIM is also included in the table.

As it can be seen from these results, both methods are not weighing the distortions equally, i.e. each one *stresses* different concepts. MSSIM hardly interprets blurring like a distortion, whereas QILV gives a high penalty to it, the lower the index value the greater the blur. On the other hand, QILV is not so affected by noise, and should be decreasing with respect to an increasing noise level. The addition of a constant is not seen at all as a source of degradation.

The first term in eq. ([18]) is the most sensitive to noise; as seen in Fig. 3, the effect of certain kinds of noise is to shift the mode of the distribution. In the QILV index, this displacement is taken into account by the mean of the local

Distortion	MSSIM	QILV	QILV$^+$
Blur 5×5	0.96	0.42	0.42
Blur 21×21	0.87	0.01	0.01
White Noise	0.63	0.92	0.01
Constant	0.86	1	1

TABLE II. Quality assessment for the black square experiment in Fig. 1.

variance. However, when dealing with a noise removal algorithm we may need an index still more sensitive to noise variations. To that end, we define a new term based on the median of the distribution, which is defined as

$$m_{V_I} = \underset{\mathbf{x}}{\text{median}}\{\hat{\sigma}_I^2(\mathbf{x})\} \quad [21]$$

Accordingly we define the new index to be

$$\text{QILV}^+(I,J) \quad = \quad \text{QILV}(I,J) \cdot \left[\frac{2m_{V_I}m_{V_J}}{m_{V_I}^2 + m_{V_J}^2}\right]^{\phi} \quad [22]$$

The black box experiment is repeated for this index and the results are shown on the last column in Table II. The value for all the distortions but the one due to noise are similar to those of QILV. The index for the white noise reflects sharply the noise content. As we will show later, such a reaction is not so clear when dealing with more complex images.

EXPERIMENTAL RESULTS

To better understand the behavior of the proposed index, a set of experiments have been carried out using different distortion sources to cover a wide range of signal transformations and corruptions. The distortions are classified as follows:

1) Linear filtering: (a) Low-pass filtering: Gaussian blur and local averaging. (b) High-pass filtering: Separable high-pass Butterworth filter. Different cut-off frequencies will be used to study the behavior at different bandpass levels.
2) Non-linear filtering: Median filtering
3) Histogram transformations: Linear ramp, gamma correction and histogram equalization.
4) Noise: (a) I.i.d. noise: Additive Gaussian, Rician, Salt and Pepper. (b) Spatially correlated: Additive Gaussian. (c) Multiplicative noise: Speckle.
5) Signal Compression: (a) DCT-based compression: JPEG. (b) Wavelet-based compression: JPEG2000. (c) Spectral-based compression: Singular Value Decomposition (SVD) (28).

Except for the histogram equalization, the degree of the distortions can be controlled by, at least, one parameter (for example the standard deviation σ for Gaussian blurring). We will study how the different quality indexes vary with the degree of distortions.

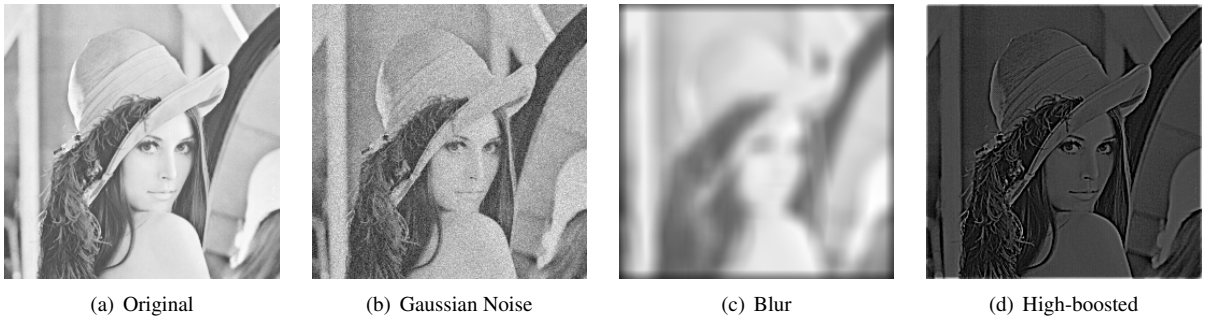
The local variances for SSIM, QILV and QILV$^+$ have been computed using a weighted moments estimator, (see eq. [12]) with a Gaussian function with $\sigma = 1.5$ and $N = 11$. Throughout our experiments, the following parameters has been used for each method: $\alpha = \beta = \gamma = 1$, for SSIM and QILV, and $\phi = 1$ for QILV$^+$; $C_1 = 6.5025$, $C_2 = 58.5225$ and $C_3 = \frac{C_2}{2}$ for SSIM –the values proposed in (2)– and $C_4 = 6.5025$, $C_5 = 58.5225$ and $C_6 = \frac{C_5}{2}$ for QILV.

Fixed-value analysis

This first experiment compares the variation of QILV and QILV$^+$ when different distortions corresponding to a fixed SSIM value are applied, and vice versa. As an illustration, different distortions are made over the image in fig. 4-(a), keeping the SSIM index constant and equal to 0.50. Parameters of each specific distortion have been manually adjusted to have the same SSIM value, namely:

1) Additive Gaussian noise (I_N), with 0 mean and $\sigma = 16$, Fig. 4-(b).

2) Local Averaging using a 17×17 kernel (I_B), Fig. 4-(c).
3) High pass version is obtained by high-boosting the image; $I_H = 1.243I - I_L$ being I the original image and I_L a low pass version obtained by the convolution of I with a 5×5 averaging kernel, Fig. 4-(d).
4) SVD is done (I_S) and the most significant eigenimage is taken, Fig. 4-(e).
5) JPEG compression (I_J) with $N = 1$ coefficient of the DCT, Fig. 4-(f).
6) JPEG 2000 compression (I_{2K}) using Jasper (29) with quality 0.005, Fig. 4-(g).



(a) Original (b) Gaussian Noise (c) Blur (d) High-boosted

(e) SVD (f) JPEG (g) JPEG2000

Figure 4. Images constructed to have the same MSSIM=0.50. Note that images with the same quality measure may have very different visual quality. The new index (QILV) seems to be more balanced.

In the following table the MSSIM is compared to the weighted SSIM, the QILV and the QILV$^+$:

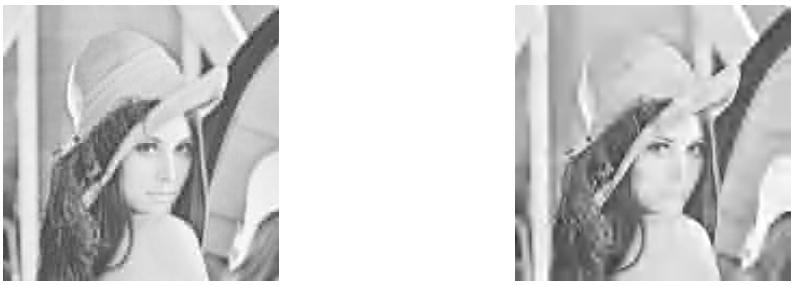
	MSSIM	WSSIM	QILV	QILV$^+$
I_N	0.50	0.73	0.89	0.21
I_B	0.50	0.30	0.01	0.005
I_H	0.50	0.49	0.83	0.79
I_S	0.50	0.26	0.004	0.0013
I_J	0.50	0.42	0.38	0.08
I_{2K}	0.50	0.35	0.03	0.02

The results for the SVD decomposition clearly reveal a situation in which SSIM index does not yield a satisfactory result. Visually, I_S –Fig. 4-(e)– is probably the *most different* image achieved for this constant SSIM index. Surprisingly, the index indicates a strong similarity although they are visually very different. Indeed, there might be an infinite number of totally different images that share the same basis. Since QILV and QILV$^+$ are based on variance distributions, this problem is detected and indicated with a low quality value.

Conversely, we now set QILV to 0.77. The distortions and parameters that achieve this fixed QILV value are:

1) Additive gaussian noise(I_N) with 0 mean and $\sigma = 18.4$, Fig. 5-(a).
2) Image blurring with a 3×3 averaging kernel (I_B), Fig. 5-(b).
3) High pass filtering computed as $I_H = 1.16I - I_L$ being I the original image and I_L a low pass version obtained by the convolution of I with a 3×3 kernel, Fig. 5-(c).
4) Image compression using the 27 most significant eigenimages computed via SVD (I_S), Fig. 5-(d).

(a) Gaussian Noise (b) Blur (c) High-boosted (d) SVD



(e) JPEG (f) JPEG2000

Figure 5. Images with QILV=0.77.

(a) Noise (b) BLUR (c) JPEG2000

Figure 6. Images with QILV$^+$=0.77.

5) And finally, image compression using JPEG (I_J) with 7 DCT coefficients , Fig. 5-(e) and JPEG 2000 (I_{2K}) with quality parameter 0.02, Fig. 5-(f).

The results in this case are as follows:

	MSSIM	WSSIM	QILV	QILV$^+$
I_N	0.45	0.68	0.77	0.16
I_B	0.88	0.89	0.77	0.63
I_H	0.40	0.38	0.77	0.77
I_P	0.78	0.84	0.77	0.77
I_J	0.81	0.81	0.77	0.74
I_{2K}	0.75	0.79	0.77	0.62

Once again, SSIM weighs more heavily noise over blur. As for QILV, the high boost case is more optimistic than what it probably should, since structural content is enhanced at the expense of removing background information. Note that the QILV$^+$ has almost the same results than QILV, but blur, JPEG2000, and most notably noise. As a point of comparison between QILV and QILV$^+$, the images that correspond to a fixed QILV$^+$=0.77 for these three cases

are on Fig. 6. We can see how noise is the only factor that clearly reveals differences between QILV and QILV$^+$. Generally speaking, QILV$^+$ is more sensitive to small noise variations than QILV, as intended.

Comparative Evaluation

Figure 7. Small subset of the images used for the comparitive study. All images are 256×256 with 256 gray levels.

In order to illustrate the behavior of the proposed indexes and to compare them with some of the state-of-the-art methods for a wide range of values, a final experiment has been carried out. 40 different images from different sources have been selected, 37 from the miscellaneous USC-SIPI Image Database[2] and other 3 as shown in the upper row of Fig. 7, all of them with size 256×256 and 256 gray levels. The images are degraded from different sources:

1) Additive Gaussian noise with zero mean and different values of σ.
2) Rician Noise with different values of σ.
3) Salt and Pepper Noise with different density levels.
4) Multiplicative Gaussian noise with different σ values.
5) Gaussian smoothing using kernels with different σ.
6) Spatially correlated additive Gaussian noise.
7) Local averaging using a $N \times N$ square window with different N values.
8) JPEG compression, taking a different number N of coefficients of the DCT in each block.
9) Singular value decomposition of the image and different number N of eigenimages considered.
10) JPEG2000 compression using different quality rates.
11) Linear correction using different slopes and intercepts.
12) Gamma correction of the image with different γ values.
13) Median filter with different window sizes.

Results for different bounded quality assessment methods are on Fig. 8 and Fig. 9. The methods selected are: SSIM (blue line), QILV (red line), QILV$^+$ (dashed red line), PQS divided by 5.797 (green line) and the M3 fuzzy measure previously defined applied over the histogram (black line). The results depicted are the average value —Fig. 8— and the standard deviation —Fig. 9— of each index along the 40 images for each value.

Note that all the indexes are monotonically decreasing with the degradation. PQS shows a good behavior in JPEG and SVD experiments, as expected, even though not clearly better than others. Additionally, when other kinds of

[2] `http://sipi.usc.edu/database/`.

distortions are presented, the index quickly falls to 0 values. SSIM as previously seen, is the index with the higher values for smoothing; however, it hardly detects a blur of the edges as a distortion. Results for the compression analysis –JPEG, JPEG2000, SVD– clearly shows one of the weak points of SSIM index. The minumum posible value, that is, for the maximum degradation, is in all cases around 0.5. This drastically reduces the the numerical range of this index for compression evaluation. It would be desirable to have a behavior similar to the other indexes.

A final experiment is carried out to assess the variance distribution of each of the quality indexes for a family of images regardless of the distortion. It is reasonable to assume that a robust quality index should yield the same value when different images undergo the same distortion, as long as the images share a similar statistical behavior, i.e. the images have similar structural content. Based on this assumption, we tested the robustness of the quality indexes under consideration using the previously mentioned distortions. For each method, distortion type and strength, the sampled variance of the quality indexes was computed along the axis that contains each image realization. Each distortion is typically parameterized by one parameter, for example the degree (*strength*) of blurring, and the parameter is different in nature between distortions. To avoid a bias to any given distortions, 1000 random instances of the resulting quality index variance were uniformly drawn from the pool of distortion strengths for any given distortion type. The resulting set was 1000×15 samples of the variance distribution for each method. The distribution of variances for each index is shown in Fig. 10 using a box-plot graph (30). Both QILV and SSIM show the lowest median variance suggesting that the methods are consistent for our range of images. Quality indexes that are designed to assess a given distortion type, like QILV$^+$ and PQS, do not necessarily exhibit a low and narrow distribution of variances as shown in Fig. 10 For these cases, this experiment is not as informative and no conclusions should be drawn. M3 shows a very narrow distribution of variances with a clear bias suggesting that the method highly depends on the specific image to assess its quality.

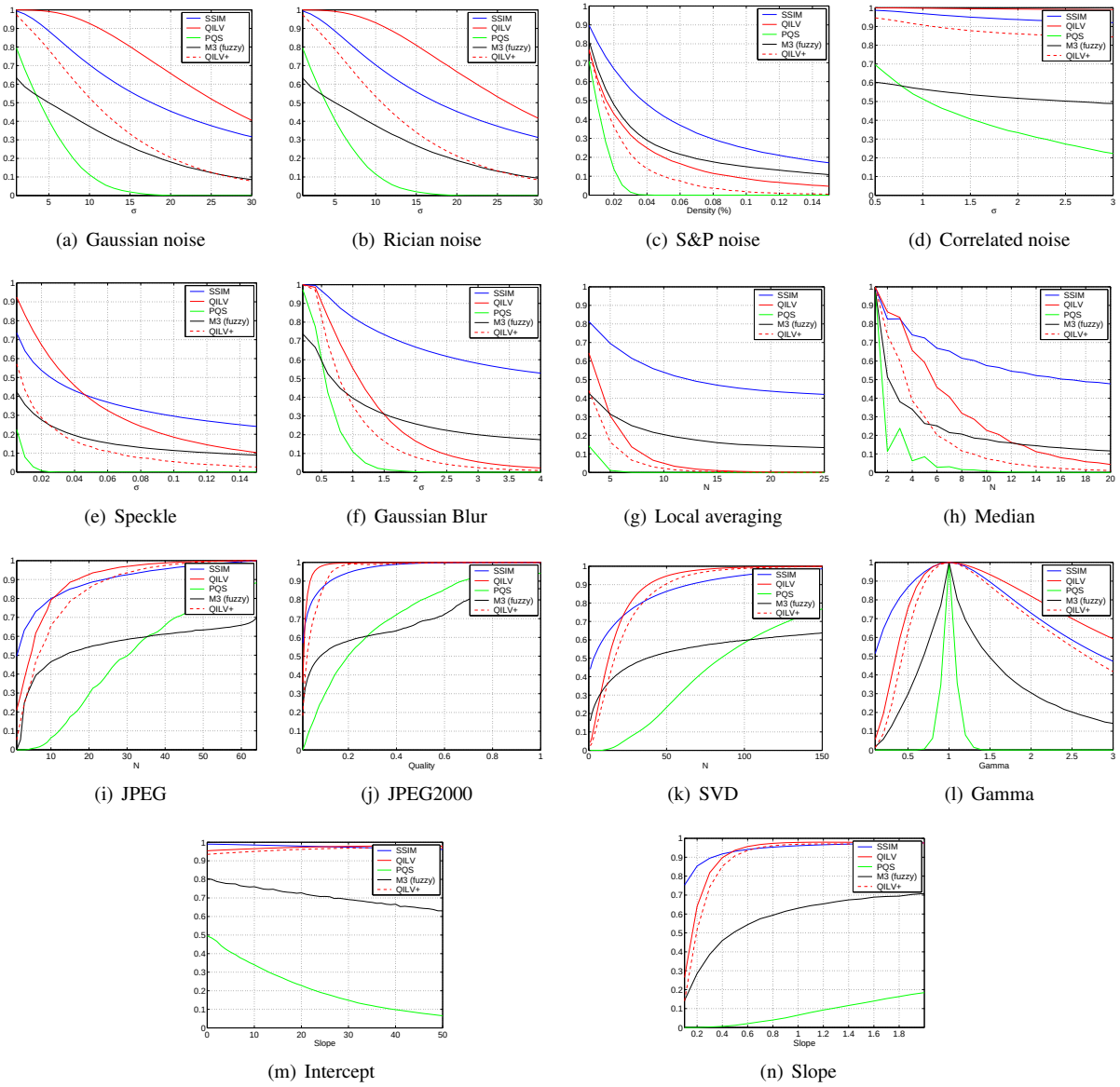
DISCUSSION

A good measure to assess the quality of an image must depend on the image structure. The way a particular distortion affects an image is related to the structure of the image itself. For instance, the more edges in the image, the more the structure is damaged by a blur, and accordingly the lower a quality index should be. So, even though it is desirable a similar behavior of an index for similar distortions, in the end the index will be ultimately related to the image itself. On the other hand, as it has been shown in the experiments, different indexes take into account different features of the image to assess its quality. It is important to know the behavior of a particular index with different distortions in order to properly understand the results. As it was shown in Figs. 8 and 9 every method has its own dynamic range for each source of degradation.

In addition, these measures must be seen as *companions* rather than *competitors*. Some of them are giving important information that cannot be left aside. In order to assess the quality of an image, more than a single index should be taken into account. When testing a noise reduction algorithm test, for instance, a good combination could be MSE, MSSIM and QILV; MSE gives an absolute error measure, MSSIM gives an idea of the noise reduction, while a low value of QILV would indicate some blurring. This multiple measures can be comprehensively presented using, for instance, Hosaka plots, as in Fig 11. It is not the scope of this paper to study how the indexes can be combined to define new ones. It is, however, apparent to us that a multidimensional representation might have to be chosen to be able to capture the information conveyed by each index.

CONCLUSIONS

A new method for image quality assessment has been introduced; it is based on the distribution of the local variance of the data. According to some previous studies, the local variance distribution gives accurate information about the structure of an image. The effect of some distortions over such distribution is well known, which makes it a suitable measure for a structural quality index to rely on. From the experiments carried out in this paper it is our understanding that the quality indexes derived from a measurement of the local variance distribution correspond more closely to those expected from subjective visual assessment (concerning structural information) than methods previously reported. A second index, namely QILV$^+$ has also been introduced to be used in noisy environments. The joint use of the indices here proposed with others previously described in the literature is a field that should be further explored.



(a) Gaussian noise (b) Rician noise (c) S&P noise (d) Correlated noise

(e) Speckle (f) Gaussian Blur (g) Local averaging (h) Median

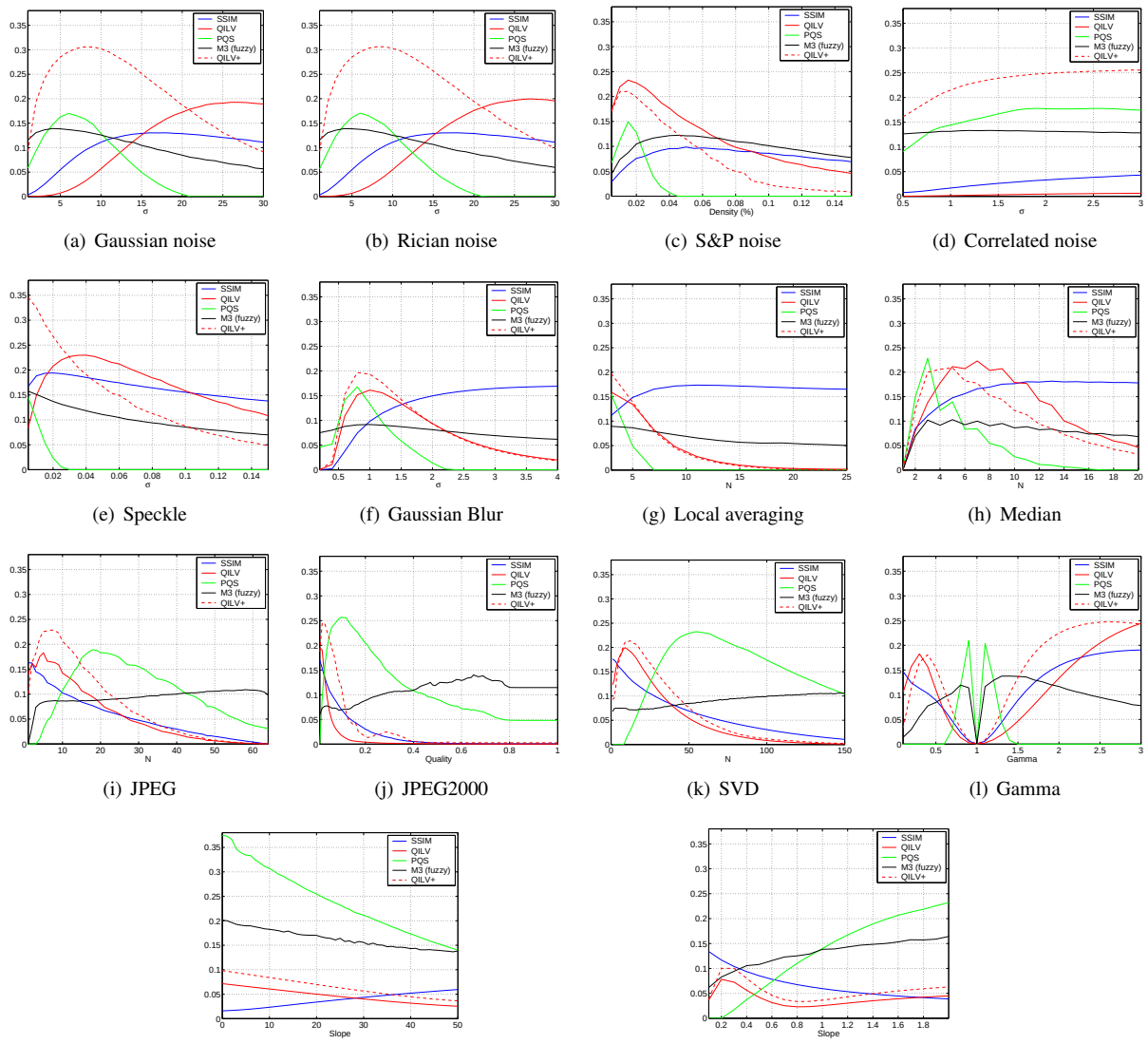
(i) JPEG (j) JPEG2000 (k) SVD (l) Gamma

(m) Intercept (n) Slope

Figure 8. AVERAGE: Evolution of different bounded quality measures: MSSIM (blue), QILV (red), PQS (green), M3 (black) QILV$^+$ (dahed red).

REFERENCES

1. J. S. Lim, Two Dimensional Signal and Image Processing, Prentice Hall, Englewood Cliffs, NJ, 1990.
2. Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Proccess. 13 (4) (2004) 600–612.
3. H. Tang, L. Cahill, A new criterion for the evaluation of image restoration quality, in: Proc. of the IEEE Region 10 Conf. Tencon 92, Melbourne, Australia, 1992, pp. 573–577.
4. A. Eskicioglu, P. Fisher, Image quality measures and their performance, IEEE Tr. Comm. 43 (12) (1995) 2959–2965.
5. X. Li, Blind image quality assessment, in: Proc. of the IEEE Int. Conf. Im. Proc, ICIP, Vol. I, Rochester, NY, 2002, pp. 449–452.
6. Z. Wang, H. Sheikh, A. Bovik, No-reference perceptual quality assessment of JPEG compressed images, in: Proc. of Int. Conf. on Image Processing, Vol. 1, 2002, pp. I–477 – I–480.
7. H. Sheikh, A. Bovik, L. Cormack, No-reference quality assessment using natural scene statistics: JPEG2000, IEEE Trans. Image Proccess. 14 (11) (2005) 1918–1927.
8. M. Kupinski, J. Hoppin, E. Clarkson, H. Barrett, G. Kastis, Estimation in medical imaging without a gold standard, Acad Radiol 9 (3) (2002) 290–297.



(a) Gaussian noise (b) Rician noise (c) S&P noise (d) Correlated noise

(e) Speckle (f) Gaussian Blur (g) Local averaging (h) Median

(i) JPEG (j) JPEG2000 (k) SVD (l) Gamma

(m) Intercept (n) Slope

Figure 9. STD: Evolution of different bounded quality measures: MSSIM (blue), QILV (red), PQS (green), M3 (black) QILV$^+$ (dahed red).

9. S. Aja-Fernández, R. San-José-Estépar, C. Alberola-López, C.-F. Westin, Image quality assesment based on local variance, in: Proc of the 28th EMBC, New York, 2006, pp. 4815–4818.
10. B. Girod, Digital images and human vision, MIT Press, Cambridge, MA, USA, 1993, Ch. What's wrong with mean-squared error?, pp. 207–220.
11. A. Przelaskowski, Vector quality measure of lossy compressed medical images, Computers in Biology and Medicine 34 (3) (2004) 193–207.
12. P. Cosman, R. Gray, R. Olshen, Evaluating quality of compressed medical images: SNR, subjective rating, and diagnostic accuracy, Proceedings of the IEEE 82 (6) (1994) 919–932.
13. N. Damera-Venkata, W. Kite, T.D.; Geisler, B. Evans, A. Bovik, Image quality assessment based on a degradation model, IEEE Trans. Image Proccess. 9 (4) (2000) 636–650.
14. M. Miyahara, K. Kotani, V. R. Algazi, Objective picture quality scale (PQS) for image coding, IEEE Tr. Comm. 46 (9) (1998) 1215–1226.
15. H. Sheikh, M. Sabir, A. Bovik, A statistical evaluation of recent full reference image quality assessment algorithms, IEEE Trans. Image Proccess. 15 (11) (2006) 3440–3451.
16. H. Sheikh, A. C. Bovik, G. de Veciana, An information fidelity criterion for image quality assessment using natural scene statistics, IEEE Trans. Image Proccess. 14 (12) (2005) 2117–2128.
17. H. Sheikh, A. Bovik, Image information and visual quality, IEEE Trans. Image Proccess. 15 (2) (2006) 430–444.

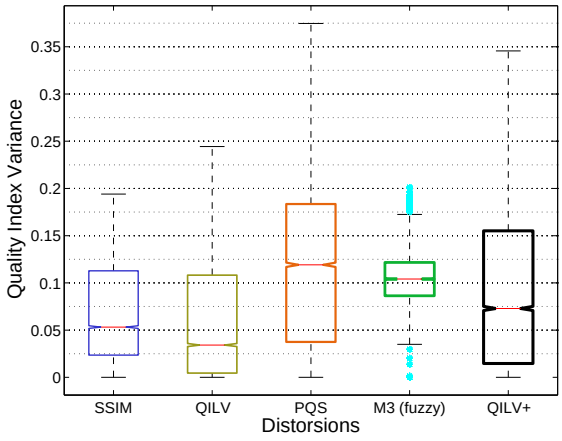
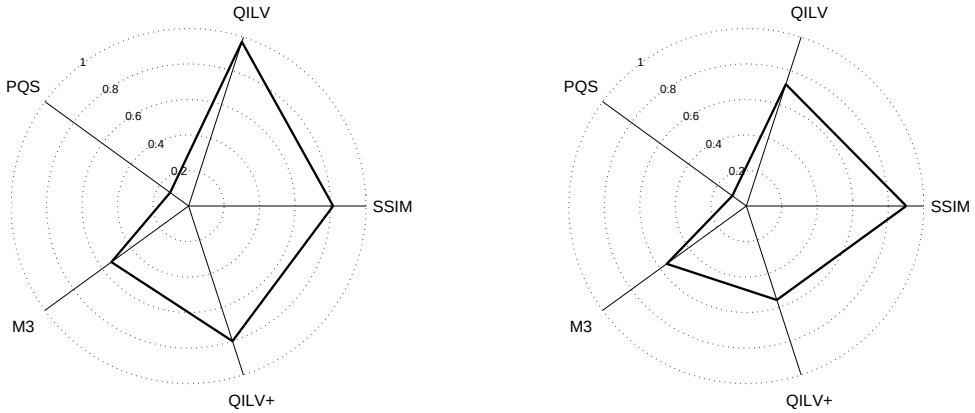


Figure 10. Quality Indexes variance distribution



(a) Gaussian Noise (b) Gaussian Blur

Figure 11. Hosaka plot for two different distortions over the second image in Fig. 7. (a) Gaussian Blur with σ_n=7. (b) Gaussian blur with $\sigma = 1$.

18. Z. Wang, A. C. Bovik, A universal image quality index, IEEE Signal Processing Letters 9 (3) (2002) 81–84.
19. Z. Wang, A. C. Bovik, E. P. Simoncelli, Handbook of image and video processing, 2nd ed., Academic Press, 2005, Ch. Structural approaches to image quality assessment.
20. D. V. der Weken, M. Nachtegael, E. Kerre, Using similarity measures and homogeneity for the comparison of images, Image and Vision Computing 22 (9) (2004) 695–702.
21. D. V. der Weken, M. Nachtegael, V. D. Witte, S. Schulte, E. Kerre, A survey on the use and the construction of fuzzy similarity measures in image processing, in: Proc. of IEEE CIMSA 2005, 2005, pp. 187–192.
22. D. V. der Weken, M. Nachtegael, E. Kerre, Combining neighbourhood-based and histogram similarity measures for the design of image quality measures, Image and Vision Computing 25 (2) (2007) 184–195.
23. D. V. der Weken, M. Nachtegael, E. Kerre, An overview of similarity measures for images, in: Proc. of ICASSP '02, Vol. 4, 2002, pp. IV–3317–IV–3320.
24. L. Sendur, I. Selesnick, Bivariate shrinkage with local variance estimation, IEEE Signal Processing Letters 9 (12) (2002) 438–441.
25. S. Aja-Fernández, C. Alberola-López, On the estimation of the coefficient of variation for anisotropic diffusion speckle filtering, IEEE Trans. Image Proccess. 15 (9) (2006) 2694–2701.
26. S. Aja-Fernández, G. Vegas-Sánchez-Ferrero, M. Martín-Fernández, C. Alberola-López, Automatic noise estimation in images using local statistics. Additive and multiplicative cases, Image Vis. Comput.In press.
27. S. Aja-Fernández, C. Alberola-López, C.-F. Westin, Noise and signal estimation in magnitude MRI and Rician distributed images: A LMMSE approach, IEEE Trans. Image Proccess. 17 (8) (2008) 1383–1398.
28. J.-F. Yang, C.-L. Lu, Combined techniques of singular value decomposition and vector quantization for image coding, IEEE Trans. Image Proccess. 4 (8) (1995) 1141–1146.
29. M. Adams, F. Kossentini, JasPer: a software-based JPEG-2000 codec implementation, in: Proc. of ICIP, Vol. 2, 2000, pp. 53–56.
30. J. W. Tukey, Exploratory Data Analysis, Addison-Wesley, Reading, MA, 1977.