Proceedings of the 28th IEEE
EMBS Annual International Conference
New York City, USA, Aug 30-Sept 3, 2006

Image Quality Assessment based on Local Variance

Santiago Aja-Fernández, Raúl San José Estépar, Carlos Alberola-López and Carl-Fredrik Westin

Abstract— **A new and complementary method to assess image quality is presented. It is based on the comparison 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. We show that this new index outperforms other methods for the assessment of image quality in medical images.**

I. INTRODUCTION

It is well known that images can suffer distortion due to several sources, from the acquisition process itself to compression, noisy channels and so on. On the other hand, images can also undergo quality improvement processes, like enhancement or restoration techniques [1]. In every case it is useful to quantify the *quality* of such 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 pixel-to-pixel error measurement, like MSE [3], [4] and other error measurements [5]. Alternatively Wang *et al.* [2] proposed the Structural Similarity (SSIM) index. This method, based on the structural information of the image, has proved to be a good measure for very different kinds of images, from natural scenes to medical images [6]. However, one may think of situations in which the information provided by this index does not match a subjective quality judgement. It is due to the bias each method has towards the *image statistic* it is using to measure. Some other quality assessment methods based on different features may give more accurate information of the global quality.

A global quality metric is a ubiquitous problem in the processing of medical images. The structural content provided by scanning devices cannot be compromised by filtering methods if the result is intended to be fed into the clinical work-flow. Being able to objectively quantify the quality gain with respect to the originally scanned image, as well as the quality in the adquisition is crucial for the adoption of processing techniques without compromising the diagnostic value. In this paper, we present a method based on the distribution of the local variance in the images; with the aim to better handle the non-stationarity of the images to be compared. Non-stationary processes naturally arise on images where structures are present. Changes on the structural behavior will lead to a change of the non-stationarity behavior. The new method can be seen as a stand alone new index, or as a complement to other existing methods, such as the SSIM.

The authors acknowledge CICyT for research grants TIC2001-3808-C02-02 and TEC2004-06647-C03-01, FIS for grants G03/135 and PIO-41483, Junta de Castilla y León for grant VA075A05, the European Com. for the funds associated to the NoE SIMILAR (FP6-507609) and the MEC/Fulbright Comission for research grant FU2005-0716.

S. Aja-Fernández and C. Alberola-López are with the LPI, ETSI Telecomunicación Universidad de Valladolid.

S. Aja-Fernández, R. San José Estépar and C.F. Westin are with Brigham and Women's Hospital, Harvard University.

II. BACKGROUND

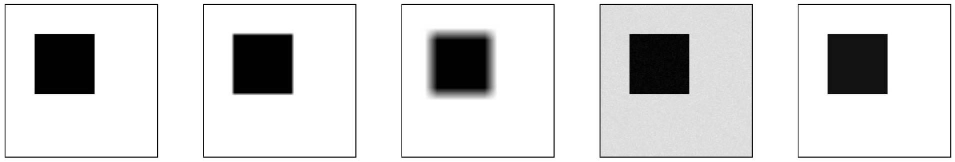
One of the most used methods to quantify the quality of an image is the Mean Square Error (MSE) [3], [4]. It gives a measure of how pixelwise similar two images are. Though, it does not take into account any structural information of them. Alternatively, some other methods have been proposed into the Medical Image field [5]. The limitations of such methods have been widely reported in literature.

In [2] Wang *et al.* proposed a new quality (full-reference) assessment method based on the structural similarity of two images I and J. Up to date, this method has proved to be versatile and robust in many different environments. It uses three levels of comparison[1]:

1) Luminance comparison:
$$l(I,J)=\frac{2\mu_I\mu_J+C_1}{\mu_I^2+\mu_J^2+C_1}$$
with μ_I and μ_J the local mean of the images I and J, and C_1 a constant.
2) Contrast comparison:
$$c(I,J)=\frac{2\sigma_I\sigma_J+C_2}{\sigma_I^2+\sigma_J^2+C_2}$$
with σ_I and σ_J the local standard deviation of the images I and J, and C_2 a constant.
3) Structure comparison:
$$s(I,J)=\frac{\sigma_{IJ}+C_3}{\sigma_I\sigma_J+C_3}$$
with σ_{IJ} the local correlation coefficient between the images I and J, and C_3 a constant.

The local SSIM index is defined as
$$\text{SSIM}(I,J)=[l(I,J)]^\alpha\cdot[c(I,J)]^\beta\cdot[s(I,J)]^\gamma \quad (1)$$
and with a proper parameter election [2] it becomes
$$\text{SSIM}(I,J)=\frac{(2\mu_I\mu_J+C_1)(2\sigma_{IJ}+C_2)}{(\mu_I^2+\mu_J^2+C_1)(\sigma_I^2+\sigma_J^2+C_2)} \quad (2)$$
The overall value is obtained using the mean of the local SSIM (with acronym MSSIM): Some variations of the original methods have been proposed elsewhere, like using a weighted sum instead of the mean [7].

[1] μ_I and σ_I in this section must be understood to be local, i.e. $\mu_{I_{i,j}}$ and $\sigma_{I_{i,j}}$.



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

Fig. 1. Black square (256 gray levels). (a) Original Image. (b) Blurred image using a square 5×5 window (MSSIM=0.9637). (c) Blurred Image using a square 21×21 window (MSSIM=0.8689) (d) Image with additive Gaussian noise 0 mean and $\sigma = 5/255$ (MSSIM=0.6278) (e) Image plus constant 10 (MSSIM=0.8526).

III. IMAGE QUALITY ASSESSMENT BASED ON LOCAL VARIANCE

A. *Some previous experiments*

Although the SSIM index has shown to be a very useful index in many experiments, cases may arise in which the quality measure obtained does not match properly a subjective judgment based on the visual information. As an example, consider the image in Fig. 1-(a). Some distortions are inserted on it:

- The image is blurred via convolution with a 5×5 averaging kernel, Fig. 1-(b) The resulting image has a MSSIM=0.9637.
- The image is blurred via convolution with a 21×21 averaging kernel, Fig. 1-(c) The resulting image has a MSSIM=0.8689.
- The image is corrupted by Gaussian noise with 0 mean and $\sigma = 5/255$. The resulting image has a MSSIM=0.6278.
- Finally, a constant (10) is added to the image, resulting MSSIM=0.8526.

From these examples it is easy to see that the 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 minimally taken as a degradation, although for medical images it may constitute an important structural change; on the other hand white noise is seen as a substantial degrading effect, when in fact the structures may be *clearer* to the human eye than the blurred ones. Some other related examples will be shown in section IV. In order to reduce this bias alternative quality measures should be conceived, and they should rely on different structural information. In next section a new such method is introduced.

B. *Quality Index based on Local Variance*

The new index we propose is based on the assumption that a great amount of the structural information of an image is coded in its local variance distribution. The SSIM index, for example, calculates the local variances of both images, but the global index takes into account only the mean of those values. Thus, the non-stationarity of the image is ignored. A further comparison based on the local variances features can help us properly compare two images.

The local variance of an image I is defined as $\mathrm{Var}(I_{i,j}) = E\left\{(I_{i,j} - \bar{I}_{i,j})^2\right\}$, being $\bar{I}_{i,j} = E\left\{I_{i,j}\right\}$ the local mean of the image. It may be estimated using a weighted neighborhood $\eta_{i,j}$ (such as Gaussian functions [2]) centered about the pixel under analysis with respective weights ω_p, as[2]

$$\mathrm{Var}(I_{i,j}) = \frac{\sum\limits_{p \in \eta_{i,j}} \omega_p \left(I_p - \bar{I}_{i,j}\right)^2}{\sum\limits_{p \in \eta_{i,j}} \omega_p} \tag{3}$$

with

$$\bar{I}_{i,j} = \frac{\sum\limits_{p \in \eta_{i,j}} \omega_p I_p}{\sum\limits_{p \in \eta_{i,j}} \omega_p}. \tag{4}$$

The size of the neighborhood $\eta_{i,j}$ should be related to the scale of the image structures expected for a particular application. The estimated local-variance of the image will be used as a quality measure of the structural similarity between two images. In fact, we will use some of its statistics. First, the mean of the local variance μ_{V_I} is estimated as

$$\hat{\mu}_{V_I} = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} \mathrm{Var}(I_{i,j}) \tag{5}$$

The (global) standard deviation of the local variance is defined as

$$\sigma_{V_I} = \left(E\left\{(\mathrm{Var}(I_{i,j}) - \mu_{V_I})^2\right\}\right)^{1/2}$$

and it can estimated as

$$\hat{\sigma}_{V_I} = \left(\frac{1}{MN-1} \sum_{i=1}^{M} \sum_{j=1}^{N} \left(\mathrm{Var}(I_{i,j}) - \mu_{V_I}\right)^2\right)^{1/2} \tag{6}$$

Finally, the covariance between the variances of two images I and J is defined as

$$\sigma_{V_I V_J} = E\left\{(\mathrm{Var}(I_{i,j}) - \mu_{V_I})(\mathrm{Var}(J_{i,j}) - \mu_{V_J})\right\} \tag{7}$$

and its estimator is

$$\hat{\sigma}_{V_I V_J} = \frac{1}{MN-1} \sum_{i=1}^{M} \sum_{j=1}^{N} \left(\mathrm{Var}(I_{i,j}) - \mu_{V_I}\right)\left(\mathrm{Var}(J_{i,j}) - \mu_{V_J}\right) \tag{8}$$

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

$$\mathrm{QILV}(I,J) = \frac{2\mu_{V_I}\mu_{V_J}}{\mu_{V_I}^2 + \mu_{V_J}^2} \cdot \frac{2\sigma_{V_I}\sigma_{V_J}}{\sigma_{V_I}^2 + \sigma_{V_J}^2} \cdot \frac{\sigma_{V_I V_J}}{\sigma_{V_I}\sigma_{V_J}} \tag{9}$$

Note that though there is a great (intentional) similarity between eq. (9) and the SSIM index, the latter is the mean of

[2]For the sake of simplicity we will denote Var(.) both the variance and its estimator.

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

Fig. 2. (a) Original Image. Images constructed to have the same SSIM=0.50: (b) White Noise added, (c) blur distortion (d) high-boosted (e) Singular value decomposition (most significant eigenimage). Note that images with the same quality measure may have very different visual quality. The new index (QILV) seems to be more balanced.

the local statistics of the images, and the former deals with the global statistics of the local variances of the images.

The first term in eq. (9) carries out a comparison between the mean of the local variance distributions of both images. The second one compares the standard deviation of the local variances. The third term is the one to introduce spatial coherence. To avoid some computational problems with small values, some constants may be added to every term in eq. (9).

IV. SOME EXPERIMENTS

In the example of the black square (see Fig.1) the previous results are compared now with the QILV:

Distortion	MSSIM	QILV
Blur 5×5	**0.96**	0.46
Blur 21×21	0.87	0.01
White Noise	0.63	0.92
Constant	0.85	**0.98**

As it can be seen from these results, both methods are not weighting the distortions equally, i.e. each one *highlights* different directions. MSSIM hardly interprets blurring like a distortion, while QILV gives a very low value to it, the lower the value the greater the blur.

To better understand the behavior of this index we will carry out another experiment. Some distortions are made over the image in fig. 2-(a), keeping the SSIM index constant to 0.50 (parameters have been manually adjusted to have the same SSIM value):

1) White noise is added (I_N), with 0 mean and $\sigma = 16/255$, Fig. 2-(b).
2) The image is blurred using a 17×17 averaging kernel (I_B), Fig. 2-(c).
3) A high pass version is obtained as $I_H = 1.243I - I_L$ 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. 2-(d).
4) A Singular Value Decomposition of the image is done (I_S), and the most significant eigenimage is taken, Fig. 2-(e).

In the following table the MSSIM is compared to the weighted SSIM and to the QILV:

	MSSIM	WSSIM	QILV
I_N	0.50	0.73	0.83
I_B	0.50	0.30	0.01
I_H	0.50	0.49	0.89
I_S	0.50	0.26	0.004

The results for the SVD analysis clearly shows one of the weak points of SSIM index. Visually, I_S is 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. But since QILV is based on variance distributions, this problem is detected and indicated with a low value. To analyze this effect, in Fig. 3 an image is compared with itself, but only taking some of its most significant eigenimages (this number is the base variable in the plots). In both experiments, QILV initially departs from (almost) zero for one eigenimage, while the SSIM index is quite larger.

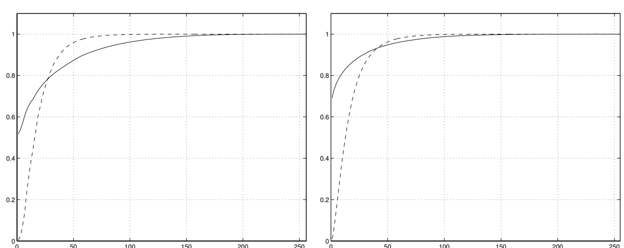


Fig. 3. Quality assessment between and image and a SVD version of itself. x-axis: number of most significant eigenimages. y-axis: SSIM index (solid) and QILV index (dashed). Left: Lenna Image. Right: Average of 29 images from LIVE database [8].

The behavior of the QILV may be understood by comparing the local variance distribution of each image, Fig. 4. The effect of blurring the image is translated into a narrowing of the distribution, i.e. the standard deviation of the variance will be much smaller in the blurred image than in the original one. On the other hand, adding white noise is equivalent to adding a constant (equal to the variance of noise) to the original distribution.

For further insight, we now set QILV to 0.77: White noise is added (I_N), with 0 mean and $\sigma = 18.4/255$, Fig. 5-(a). The image is blurred using a 3×3 averaging kernel (I_B), Fig. 5-(b). A high pass version is obtained 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). A Singular Value Decomposition of the image is done (I_S), and the 27 most significant eigenimages are taken, Fig. 5-(d). Parameters are appropriately chosen so that QILV$= 0.77$ in the four cases. Results are as follows:

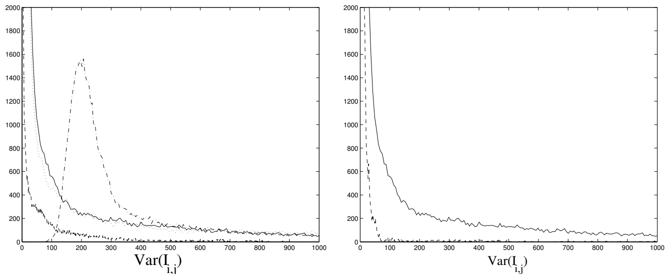


Fig. 4. Local variance distribution for Lenna, fixed SSIM=0.50. Left: Original image (solid line), noisy image (dash-dotted), blurred image (dashed) and high boosted (dotted). Right: Original image (solid line), SVD most significant eigenimage (dashed). The effect of blurring and SVD decomposition is narrowing the variance distribution, while the added Gaussian noise just change its mean.

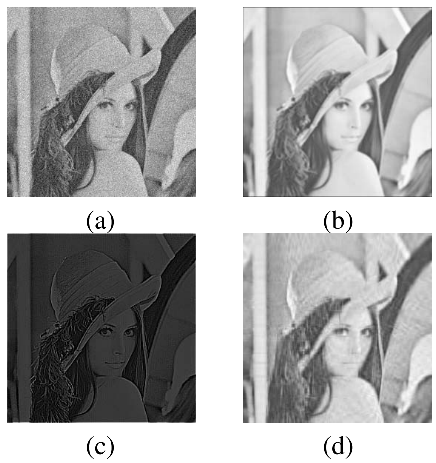


Fig. 5. Images with QILV=0.77: (a) White Noise added, (b) blur distortion (c) high-boosted (d) Singular value decomposition (27 most significant eigenimages).

	MSSIM	WSSIM	QILV
I_N	0.45	0.68	0.77
I_B	0.88	0.89	0.77
I_H	0.40	0.38	0.77
I_P	0.78	0.84	0.77

Once again, SSIM weighs noise over blur. As for QILV, the high boost case is more optimistic than what it probably sould, since structural content is enhanced at the expense of removing background information.

As a final experiment, consider the MRI in fig. 6 corrupted with Rician noise (see all the measures in the caption of the figure). A Wiener filtering or an averaging filter may be applied to reduce such noise. In both cases the SSIM index grows (with respect to the noisy case), but in the second case there is a clearer blur that affects further structural processing and interpretation of the image. On the other hand, QILV highlights this effect, which, from a clinical point of view is more relevant.

V. CONCLUSIONS

A new method for image quality assessment has been introduced; it is based on distribution of the local variance of the data. From the experiments carried out in this paper it is our understanding that the quality indices given by this method correspond more closely to those expected from subjective visual assessment (concerning structural information) than methods previously reported. In order to account for other effects not considered in the paper (or the high boost experiment indeed considered) probably a combination of SSIM and QILV may give interesting results.

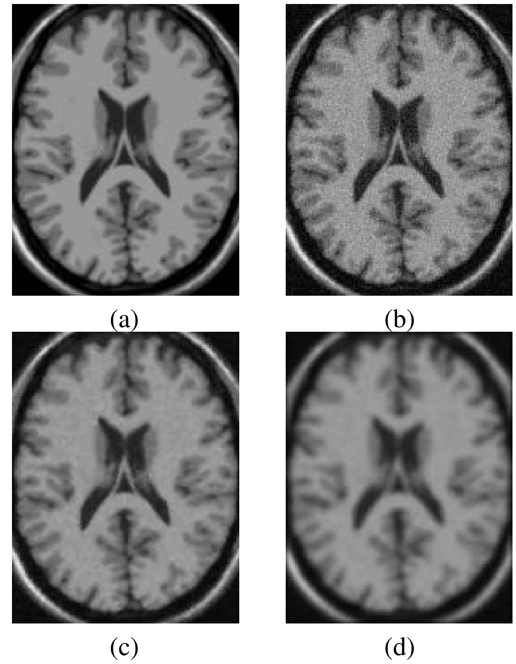


Fig. 6. Real application: MRI (a) Original Image, (b) Image with Rician Noise ($\sigma = 10/255$). $SSIM = 0.74$, $QILV = 0.98$. (c) Wiener filtering of Noisy Image. $SSIM = 0.91$, $QILV = 0.96$. (d) Averaging filter of the Noise image (using a 5×5 square window). $SSIM = 0.86$, $QILV = 0.57$. Original image from Brainweb [9].

REFERENCES

[1] J. S. Lim, *Two Dimensional Signal and Image Processing*. Englewood Cliffs, NJ: Prentice Hall, 1990.

[2] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[3] H. Tang and L. Cahill, "A new criterion for the evaluation of image restoration quality," in *Proc. of the IEEE Region 10 Conf. Tencon 92*, Melbourne, Australia, Nov. 1992, pp. 573–577.

[4] A. Eskicioglu and P. Fisher, "Image quality measures and their performance," *IEEE Tr. Comm.*, vol. 43, no. 12, pp. 2959–2965, Dec. 1995.

[5] A. Przelaskowski, "Vector quality measure of lossy compressed madical image," *COmputers in BIology and Medicine*, vol. 34, pp. 193–207, 2004.

[6] S. Aja-Fernández and C. Alberola-López, "On the estimation of the coefficient of variation for anisotropic diffusion speckle filtering," *IEEE Trans. Image Proccess.*, in press.

[7] Z. Wang, A. C. Bovik, and E. P. Simoncelli, *Handbook of image and video processing, 2nd ed.* Academic Press, 2005, ch. Structural approaches to image quality assessment.

[8] Live database. Laboratory for Image and Video Engineering (LIVE). The University of Texas, Austin. [Online]. Available: http://live.ece.utexas.edu/research/quality/subjective.htm

[9] Brainweb. McConnell Brain Imaging Centre. Montreal Neurological Institute, McGill University. [Online]. Available: http://www.bic.mni.mcgill.ca/brainweb/