An Algorithm for Quality Assessment of Images with Contrast Changes and Mean Shifts

Ivan Mochalov, Andrew Priorov, Alexandr Prozorov
P.G. Demidov Yaroslavl State University
Yaroslavl, Russia
yar_panda@yahoo.com, andcat@yandex.ru, alexprozoroff@gmail.com

Abstract—**The article presents the algorithm of an objective quality assessment of the images distorted by contrast change and mean shift. The stages of the development are considered. Full-reference image quality assessment algorithm and non-reference one are offered. The parameters of the assessment are obtained from the histograms of the original image and the distorted one. Also linear regression model is used to calculate the metric. Regression analysis methods are applied to determine effective subset of parameters.**

I. INTRODUCTION

Let us consider the problem of visual quality assessment of the images. There are several ways to do this: an assessment involving experts, the so-called subjective assessment, and an assessment using algorithms, so-called objective assessment.

In order to show the relevance of this subject, it should be noted that the automatic visual quality assessment of the images is the most demanded where necessary to optimize the image processing algorithms. For example, it is necessary for the determination of various parameters of compression algorithms, noise reduction or image normalization. Similar problems are fairly common. But the required number of visual quality assessments does not allow to involve experts for this task.

It should also be noted that there are several types of objective assessments: full-reference, no-reference and intermediate. Full-reference assessment has the highest correlation with the estimates of experts. Therefore it is better to use these methods in the task of optimization of algorithms for image processing. In the other tasks, where we haven't original image non-reference assessment should be used.

II. STATE OF THE ART

Nowadays, there are many proposed ways of full-reference image quality assessment. These methods are based on various empirical regularities of the human perception system. The basis of algorithms was created for building the basic methodology of the experiment. Let us describe this methodology to stick to it in the future. A priori there is a base of reference images and their distorted versions. Each distorted image has an average rating of experts.

On the basis of revealed regularities authors offer their methods of assessing visual quality. In order to select the parameters for new methods, it is preferable to use your own database, or use the base that differs from the test one. Further the test image base is checked for correlation of proposed algorithm's assessments and expert evaluations. Typically, this is used Pearson, Spearman and Kendall correlation coefficients. The first one is parametric and others are based on order statistics.

It is difficult to say which of the criteria is the most popular to date. Because of the calculating simplicity and historical reasons, MSE and PSNR are still in demand. More recent and breakthrough algorithm SSIM and its variety UQI with more powerful and widely used MSSSIM [1] was purposed. These methods are based on quality index at each block of image. For those 8x8 square blocks local statistics are computed. If signal in current block of original image is x, and signal of observed block is y, then $SSIM$ can be computed as follow:

$$SSIM(x,y) = [l(x,y)]^{\alpha}[c(x,y)]^{\beta}[s(x,y)]^{\gamma}, \quad (1)$$

where $l(x,y)$is luminance comparison function, defined as

$$l(x,y) = \frac{2\mu_x\mu_y+C_1}{\mu_x^2+\mu_y^2+C_1}, \quad (2)$$

$c(x,y)$ is contrast comparison function that has a similar form

$$c(x,y) = \frac{2\sigma_x\sigma_y+C_2}{\sigma_x^2+\sigma_y^2+C_2}, \quad (3)$$

and $S(x,y)$ is structure comparison function, defined as

$$s(x,y) = \frac{\sigma_{xy}+C_3}{\sigma_x\sigma_y+C_3}, \quad (4)$$

where $\alpha, \beta, \gamma, C_1, C_2, C_3$ are non-negative parameters used to adjust the relative importance of the three components (α, β, γ) and to maximize correlation with experts scores

(C_1, C_2, C_3). A number of other algorithms, for example, HQI [2] was also proposed. According to the authors, FSIM [3] is the most effective open-source algorithm. Of course, the result depends on the test base of images. In this context we note base TID2008 [4] as the most complete database of color images with a variety of types of distortions.

III. Problem Statement

As pointed out in [5], the biggest problem with the automatic image quality assessment is the processing of images distorted by contrast change or mean shift. In terms of a simplified model, these changes can be described as a change in the variance and the mean value of the image. It is not difficult to see that for such distortions image remains almost completely (up to quantization) correlated with the original image. The point is that for images, which have 8 bits per color channel, the dynamic range contains [0..255] brightness values. Thus, when changing contrast or brightness histogram of the image may be placed beyond the dynamic range and trimmed out. Thus, the Pearson correlation coefficient ceased to be equal to 1, which should also be considered.

Why do these distortions cause difficulties in the assessment. For most of the criteria there are two rules:

1) When increasing distortion factor, expert assessments are reduced.
2) When increasing distortion factor, correlation between distorted and original images are reduced.

Analysis of the results of expert assessments of images with contrast and mean distortions shows that increasing the contrast is practically not perceived as lower quality. At least as long as there is no significant change in the histogram for the reasons mentioned above. This result shown in Fig. 1 where we have dependence of mean opinion scores (MOS) on dispersion difference.

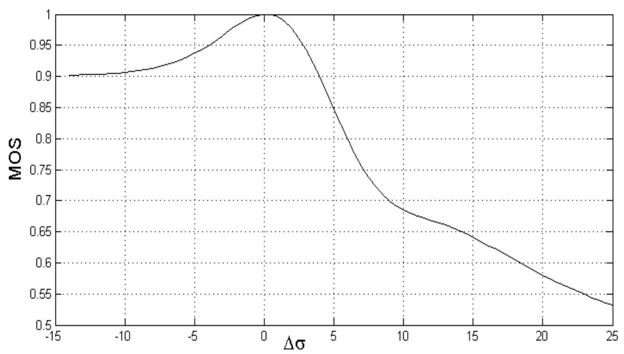

Fig. 1. Dependence of MOS on the variance change

If we consider an objective image quality assessment, as an attempt to define the algorithm that person operates when evaluating image distortions, it can be assumed that in the cases of contrast and brightness changes we have to use separate algorithms.

The aim of this work is to develop an algorithm of objective image quality assessment for the images distorted by contrast and brightness change.

Now let's concretize our task. The developed algorithm must be applied for objective evaluating the quality of the images when the correlation between distorted and reference image is close to 1. Otherwise, we will use one of the existing solutions or the algorithm that will be developed in the future. It is also possible to develop an aggregation algorithm for the case of normalization. Thus, in the first stage, if the mean and variance of the distorted image deviate from the corresponding mean and variance of the reference image, then the evaluation of the quality of the image can be performed for the case of changing the contrast and brightness. After that normalization is performed by a linear transformation of distorted image samples in order to match the mean and variance. Thus, the quality of normalized distorted image can be estimated. The last step is the aggregation of assessments of the first and second stages.

Thus, we consider in detail an algorithm for estimating the quality of image with changing the contrast and brightness. Let our algorithm will be parametric. The input parameters are the average of the initial and distorted images, as well as values characterizing the contrast of the image:

1) the difference between the brightest and darkest pixels of the image,
2) image dispersion,
3) the sum of the absolute deviations.

IV. Estimation Process

Due to the good development of methods for the determination of the linear regression coefficients, it is worth it to use a linear combination of some distortion types. However, it is clear that the relationship between the input parameters and expert estimates are not linear. Because of this, it is difficult to determine the type of nonlinearity. Therefore, for this case is supposed to use the following approach:

1) transition from the input parameters to the intermediate parameters through a set of non-linear functions.
2) construction of a linear regression model.
3) sequential exclusion of factors.

Steps 2 and 3 can be easily automated. The main difficulty lies in the 1st stage. Formally, it is a transition from the 4 parameters $\mu_e, \mu, \sigma_e, \sigma$ (means and variances of the reference image with index "e" and the distorted image without index) to the N functions $F_i(\mu_e, \mu, \sigma_e, \sigma)$. For brevity, we omit the parameters and in the future we will write simply F_i. Preferably, each of these features reflect some physical regularities in the test data. Now, it should be noted that Fechner's law is valid for the perception [6]. Under this law, the perception is changing proportional to the logarithm of the stimulus:

$$G_i = k_i * \log(F_i) + C_i. \tag{5}$$

So, if all functions G_i will have a physical meaning quantify the distortion with a certain factor definite conclusions can be made. Firstly, if we measure evaluation values in a certain range, then up to a linear transformation we can write it in the following form:

$$Q = -\sum_i w_i G_i\ , \tag{6}$$

where it is desirable to follow the rule $\forall(\mathrm{i}: \mathrm{w_i} > 0)$. Now we define G_i as a distortion measure, which is equal to 0 for the case of undistorted image, and greater than or equal to 0 for the other cases. If you require the same properties of F_i,, you will have to make changes in the formula (5), so that it becomes applicable. Since the Fechner law builds dependence of perception of the stimulus, we can put as an irritant $F_i + e_i$. In this case, the formula 1 can be rewritten as follows:

$$G_i = k_i * \log(F_i + e_i) + C_i. \tag{7}$$

Let us try to get rid of some constants. To do this, we rewrite the formula (6) with (7):

$$Q = -\sum_i w_i * k_i * \log(F_i + e_i) + C_i. \tag{8}$$

Now we can redefine $w_i \equiv w_i * k_i$ and introduce variable:

$$C_0 = -\sum_i C_i. \tag{9}$$

Then the formula (4) can be rewritten:

$$Q = C_0 - \sum_i w_i * \log(F_i + e_i). \tag{10}$$

If we introduce natural requirement that $Q = C_0$ in the case of non-distorted images, and consider that $\forall(\mathrm{i}: \mathrm{w_i} > 0, \mathrm{F_i} = 0)$, it can be written as $\forall(\mathrm{i}: \mathrm{e_i} = 1)$and:

$$G_i = \log(F_i + 1). \tag{11}$$

Thus, the formula for the resulting assessing the quality of two single-channel images can be written as:

$$Q = C_0 - \sum_i w_i * \log(F_i + 1), \tag{12}$$

where$\forall(\mathrm{i}: \mathrm{F_i} \geq 0)$for distorted images and $\forall(\mathrm{i}: \mathrm{F_i} = 0)$for undistorted.

IV. Exclusion of Factors

The main objective of the development of the algorithm is maximizing correlation between subjective expert scores and algorithm results. But there are several correlation coefficients that can be used for this purpose [7]. Classical Pearson's estimator of correlation:

$$r_P(x, y) = \frac{\sigma_{xy}}{\sigma_x \sigma_y}. \tag{13}$$

As an alternative, nonparametric measures of correlation based on rank statistics can be used. In this work we use Spearman's correlation coefficient:

$$r_S = 1 - \frac{6\sum d_i^2}{N(N^2-1)}, \tag{14}$$

where N – is signal length and d_i is the difference between the ranks of corresponding values x_i and y_i,and Kendall rank correlation:

$$r_K = \frac{n_c - n_d}{\frac{1}{2}N(N-1)}, \tag{15}$$

where n_c is a number of the samples y_i, ordered in the same way as x_i, and n_d is a number of the samples, ordered differently.

The regression model described in equation (9) can be verified as follows. Firstly, it is necessary to create a set of factors that can affect the quality metric$Q$. Let

Step 1. Definition of a set of coefficients w_i in (9), maximizing the Spearman correlation coefficient for a test base.
Step 2. Sequential testing of correlation coefficients in the case of exclusion of one of the factorsw_i.
Step 3. Exclusion of factor, removal of which will result minimal reducing of a correlation coefficient.
Step 4. Iteration of step 3 until the correlation coefficient falls less than the allowable threshold.
Step 5. Sequential Testing of excluded factors in order to find a factor, the addition of which would lead to a maximum growth of the Spearman correlation coefficient between the experts estimates $Q_e$and calculated Q.
Step 6. If the iteration limit is reached - exit from the cycle - if not, return to step 1.

Such methods are given in the literature [7]. This method has specific features. Such as the way of adding and deleting the coefficients as well as the method of determining the coefficient of correlation. From the above it follows that the main objective of this work - find the set of factors G_i, presented in the form $\log(F_i + 1)$.

Let's look at the distortion model. If we are not taking into account the quantization effects described above, distortion in contrast and mean value can be represented as follows:

$$Y = aX + b. \tag{16}$$

In this case, the correlation coefficient between these two images is equal to 1. The observations show that experts perceive a distorted image, realizing that parts are distorted almost equally. In other words F_i is a dimensionless normalized value. Normalized distortion can act as F_i:

$$\frac{|\langle X\rangle-\langle Y\rangle|}{\langle X\rangle}, \frac{|\langle X\rangle-\langle Y\rangle|}{\langle Y\rangle}, \frac{|\langle X\rangle-\langle Y\rangle|}{(\langle X\rangle+\langle Y\rangle)/2}, \frac{|\sigma_X-\sigma_Y|}{\sigma_X}, \frac{|\sigma_X-\sigma_Y|}{\sigma_Y}, \frac{|\sigma_X-\sigma_Y|}{\langle X\rangle}, \text{etc.}$$

The method of factors exclusion has shown poor results (about 0.7 by Pearson and 0.69 by Spearman), selecting factors $\frac{|\langle X\rangle-\langle Y\rangle|}{(\langle X\rangle+\langle Y\rangle)/2}$ and $\frac{|\sigma_X-\sigma_Y|}{\sigma_Y}$.

Analysis of the results has shown poor correlation of factors associated with variance and with the assessment of experts. The fact is that, as was mentioned above - an increase in contrast did not lead to such decrease in quality as the reducing of contrast. Therefore it was decided to split the factor associated with variance into 2 parts:

$$F2 = \begin{cases} \frac{\sigma_X-\sigma_Y}{\sigma_Y}, \sigma_X \geq \sigma_Y \\ 0, \sigma_X < \sigma_Y \end{cases}, F3 = \begin{cases} \frac{\sigma_Y-\sigma_X}{\sigma_Y}, \sigma_X < \sigma_Y \\ 0, \sigma_X \geq \sigma_Y \end{cases}. \quad (17)$$

It's not a big deal to show that $\frac{|\sigma_X-\sigma_Y|}{\sigma_Y} = F2 + F3$. Once the factors have been applied, the correlation coefficient became equal to 0.88 by Pearson and 0.89 by Spearman. In this case, the first factor has remained constant:

$$F1 = \frac{|\langle X\rangle-\langle Y\rangle|}{(\langle X\rangle+\langle Y\rangle)/2}. \quad (18)$$

Further, each factor F2 and F3 - factors must be complemented by the second factor with a modified denominator ($\sigma_X$and $\sqrt{\sigma_X\sigma_Y}$). The results showed that it is useful to replace the denominator factor F3, which led to increase of correlation coefficient on 0.01:

$$F3 = \begin{cases} \frac{\sigma_Y-\sigma_X}{\sigma_X}, \sigma_X < \sigma_Y \\ 0, \sigma_X \geq \sigma_Y \end{cases}. \quad (19)$$

Until now, it was a question of grayscale images. However, color images are most interesting. And in the transition from the grayscale to color images - the main issue is the choice of the color model. When working with RGB model - we have to deal with 9 factors (3 on each component). Application of the factors exclusion shows that all the factors are significant.

Due to the fact that the authors have not been able to build a large enough sample of the experts assessments (as a basis for determining the regression coefficients used CSIQ base with the number of distortion types about 450), became necessary to reduce the number of regression coefficients. For this purpose color model YCbCr was chosen. It is used in H.26x codecs. This color model helps to decorrelate color components, which makes each of the factors independent. Let's define them as *G1Y, G2Y*, *G3Y*, *G1Cb* etc. The method of exclusion of factors allowed to go directly to the 4th factors with 3 unknown variables, one of which (at *G1Y*) can be taken equal to 1 for optimization by Spearman criterion:

$$Q = C_0 - G1Y - \frac{19}{32}(G2Y + G2Cr) - \frac{5}{128}G3Cb. \quad (20)$$

Optimization by 2 unknown is well known and studied. The use of multiple iterations of the optimization with pseudo-random starting points allows eliminate descent into local extremum. Last equation corresponds to the case when the weights are optimized by the criterion of maximum correlation coefficient of Spearman.

It should be noted that the use of an increasing function does not change the correlation coefficient of Spearman, however, is able to improve the value by Pearson. Therefore it is necessary to choose an increasing function that causes the results to the scale of [0..1]. Therefore proposed to take $C_0 = 1$ and use the following function:

$$L(x) = \begin{cases} x, x > \frac{1}{2} \\ \frac{1}{2} - \frac{1}{4}\left(\frac{1}{2} - x\right), -\frac{7}{8} < x < \frac{1}{2} \\ 0, x < -7\frac{1}{2} \end{cases}. \quad (21)$$

Since $x \leq 1$ because we chose $C_0 = 1$ and x is usually not less than 1, this function for the vast majority of images will increase. Fig. 2 presents function $L(x)$.

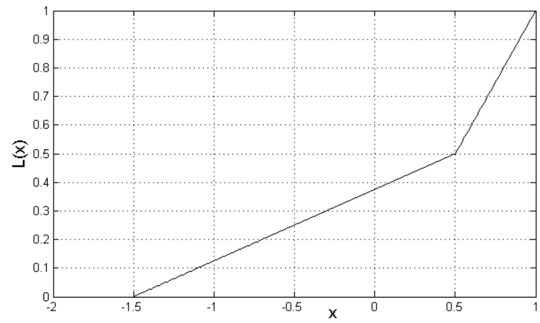

Fig. 2. L(x) function

The final criterion for assessment of the images distorted by contrast and brightness change is as follows:

$$Q = L\left(1 - G1Y - \frac{19}{32}(G2Y + G2Cr) - \frac{5}{128}G3Cb\right). \quad (22)$$

It is also possible to review the results for the extreme case of zero variance. By virtue of the fact that the image values are positive and the mean is zero as it is necessary to require zero variance. By distortion model (10) in the case ofa $= 0$, we obtain $\sigma_Y = 0$. G2Y factor in this case will not be determined and the application of the criterion (12) is unacceptable. However, assuming that $\sigma_Y \ll 1$ we can conclude,that $G2Y \gg 1$, and therefore, $Q = 0$ on the basis of (18) and (19).

V. RESULTS

The final correlation coefficients and comparison with known algorithms are given based on the database of distorted images TID-2008 for these types of distortions. For comparison, we choose the most effective algorithms, such as RIQMC [8], FSIM, HQI, MSSSIM and PSNR. We should also consider a variant of the algorithm in the absence of knowledge about the reference image. In this case, based on the distortion model (10), we can use estimates for the variance and the mean.

We assume that it is not possible to reliably estimate the average brightness of undistorted images through distorted one, and it is necessary to rely only on variance estimate. Using the rule:

$$\mu_X = \mu_Y,\ \frac{\sigma_Y}{\sigma_X} = \frac{195}{195 - offset1 - offset2}, \qquad (23)$$

where offset1 is the distance in pixels between the brightness value of 40 and a brightness threshold below which there are 2% points of the luminance histogram, and offset2 - similar distance between the brightness value of215 and a threshold, above which are 0.1% of the points of the histogram.

TABLE I. COMPARISON OF THE POPULAR APPROACHES ON TID2008 SUBSET

Criteria	Pearson correlation coefficient	Spearman correlation coefficient	Kendall correlation coefficient
PSNR	0.474	0.521	0.364
HQI	0.448	0.458	0.319
MSSSIM	0.674	0.427	0.333
IWSSIM by results of [8]	0.679	0.413	0.324
FSIM	0.497	0.357	0.253
RIQMC	0.773	0.731	N/A
PROPOSED Full-reference	**0.914**	**0.729**	**0.729**
PROPOSED Non-reference	**0.830**	**0.794**	**0.609**

As can be seen from Table I, the proposed algorithm is superior to existing counterparts. RIQMC algorithm - is the closest competitor. He does not use the knowledge that contrast change have different effects on image quality. Nevertheless, it shows good results due to the joint use of 2nd, 3rd and 4th moments. In the future, it is necessary to consider the possible revision of the proposed algorithm in this direction.

It should also be noted that non-reference algorithm has also good results. This algorithm can be used in practical tasks to assess the degree of distortion of the image and its suitability. In particular, to determine camera lens contamination automatically.

If presented metrics are effective then scatter-plots of MOS depending upon this metric should have a compact form without outliers and with a tendency to monotonous behavior. Fig. 3 shows the scatter-plot of MOS vs. PROPOSED. As it can be seen a compact region with some outliers is present. Fig. 4 presents the scatter-plot for the PROPOSED non-reference metric.

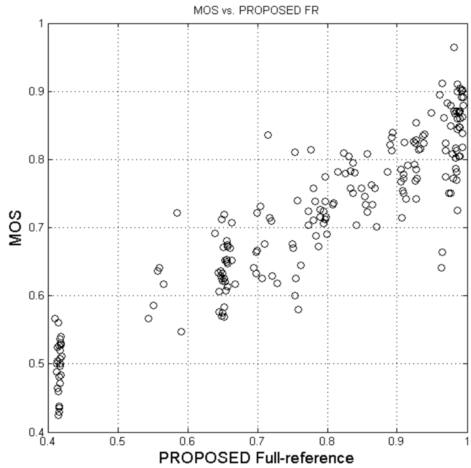

Fig. 3. MOS vs. PROPOSED Full-reference

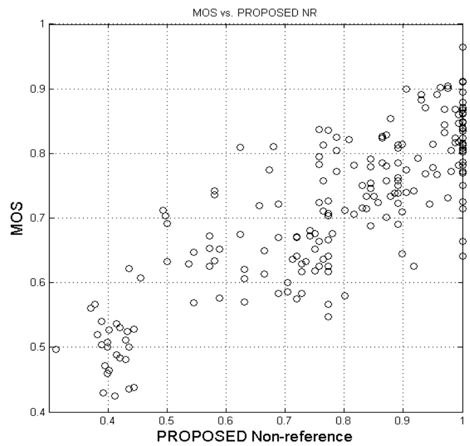

Fig. 4. MOS vs. PROPOSED Non-reference

Comparison of these plots shows that the percentage of outliers increased. This means that PROPOSED full-reference is more adequate to human perception of distorted images. It is based on usage of original image for those mean shift and contrast change is known perfectly. In general proposed algorithm can be classified as reduced-reference IQA (RR-IQA). RR method requires only a limited number of features extracted from the reference for the IQA task [9, 10]. RR-IQA is an intermediate method between FR and NR image quality assessment in terms of both correlation results and the amount of information required about original image. This type of IQA is good for tasks related to images transmission. When image A is transmitted to the client via a transmission channel with distortions. Reduced-reference factors extracted at the transmitter side are sent to the receiver. An IQA algorithm extracts factors from the received image on a client side and then calculates metric value only with pair of factors sets.

Direct application of this method hampered by the lack of problems, where considered types of distortions occur during transmission. However, the proposed method can be applied as a part of another method

VI. Conclusion

The paper presents an algorithm for objective assessment quality of the images distorted by contrast and brightness change. This algorithm shows results comparable with the algorithms known from the literature and in some cases exceeds them. Presented calculations provide an easy way to implement described algorithm.

References

[1] Z. Wang, E. P. Simoncelli and A. C. Bovik, *Multi-scale structural similarity for image quality assessment.* Invited Paper, IEEE Asilomar Conference on Signals, Systems and Computers, Nov. 2003

[2] Yalman, Y., A *Histogram based Image Quality Index.*Przeglad Elektrotechniczny, Vol. 88 (7a), pp.126-129, 2012.

[3] L. Zhang, L. Zhang, X.Mou, and D. Zhang, *FSIM: a feature similarity index for image quality assessment.* Image Processing, IEEE Trans. on *Image Proc.* V.20 , I.8 , pp. 2378-2386.

[4] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, F. Battisti, *TID2008 - A Database for Evaluation of Full-Reference Visual Quality Assessment Metrics.* Advances of Modern Radioelectronics, Vol. 10, pp. 30-45, 2009.

[5] N. Ponomarenko, O. Eremeev, V. Lukin, K. Egiazarian, M. Carli, *Modified image visual quality metrics for contrast change and mean shift accounting.* Proceedings of CADSM, pp.305-311, Polyana-Svalyava, Ukraine, 23-25 Feb., 2011.

[6] KH. Norwich, W. Wong *Unification of psychophysical phenomena: the complete form of Fechner's law*. Percept Psychophys. 1998. Jan. 60(1):174.

[7] N. Draper, H. Smith Applied Regression analysis, 3rd Edition. John Wiley & sons, Inc. 1998.

[8] KeGu, G. Zhai, X. Yang, W. Zhang, and M. Liu *Subjective and objective quality assessment for images with contrast change.* Proc. IEEE Int. Conf. Image Process 09/2013;

[9] Z. Wang and Q. Li *Information Content Weighting for Perceptual Image Quality Assessment*. IEEE Transactions on Image Processing, vol. 20, no. 5, pp. 1185–1198, May 2011.

[10] Z. Wang, H. R. Sheikh, and A. C. Bovik *Objective video quality assessment.* The Handbook of Video Databases: Design and Applications, B. Furht and O. Marques, Eds. Boca Raton, FL: CRC Press, Sep. 2003, pp. 1041–1078.