Functional quality and performance metric for some image processing applications

Chang Yun Fah, Omar Mohd Rijal and Syed Abdul Rahman Abu Bakar

***Abstract*—** One common approach to image quality and performance evaluation is to evaluate if the reference image and the processed image still remain a high level of similarity However, existing image quality metrics regarded one of the reference and processed image or both as perfect. In reality, it is not easy to obtain a perfect reference image because pre-processing procedures always generated noise into these images. In this paper, the reference and processed images are considered as imperfect and we propose a functional metric using the coefficient of determination for unreplicated linear functional relationship model as a measure of the similarity, which in turn may be used as a definition for image quality and performance indicator for some image processing applications. The sensitivity of the proposed metric is also studied and it provides a consistent interpretation for the distortion area of a processed image. Our experimental results showed that the proposed functional metric is a good quality measure for JPEG compressed image and it is sensitive in differentiating the performance or low-pass and high-pass filtered image.

***Keywords*—** functional quality metric, linear functional relationship, performance evaluation, similarity measure.

I. INTRODUCTION

DIGITAL image is the most economical and efficient medium for communicating information [1] and its applications are wide and have greatly influenced the modern lifestyle such as recording a historical event, illustrating a meaningful story or disseminating useful information. For example, a just married couple would like to capture their happy moments of marriage in pictures. While, businessmen used pictures to advertise their product. Through running a sequence of images, it becomes a movie. All these activities are widely happen in our daily life.

The demand for a higher quality image as a result of rapid progress in imaging technologies gave rise to the need to reduce intrinsic distortion of an image. In many image processing algorithms, the main objective of the algorithms is to filter the noise of a distorted image in order to obtain a processed image, which is closer to the original (reference) image. In other applications, the original image is being compressed and reconstructed (processed image) with minimum loss of information. The common interest is to study the performance of these processing algorithms in producing the desired processed image quality. One common approach to performance evaluation is to evaluate if the reference image and the processed image still remain a high level of similarity. We say that the algorithm has a good performance if the similarity value is high or vice versus. In other words, a good performance algorithm will also produce a high processed image quality. This evaluation process can be easily done visually. However, visual inspection is a highly subjective process and it is depends on the requirements of a given application [2]. Hence inconsistency results may occur because of human fatigue and subjective judgement. Consequently, many full reference objective metrics such as peak signal-to-noise ratio (PSNR) [3], root mean square error (RMSE) [4], mean structural similarity measure (MSSIM) [5], neighborhood-based similarity (S_{NBS}) [6], linear correlation coefficient (R_s^2) [7], fuzzy discrimination information measure [8], subband coded image quality control [9] and global image quality index [10] have been proposed as a quality measurement between two images, which in turn may be used as an indicator for algorithms performance evaluation.

One fundamental idea when defining the above metrics is to regard one of the original and processed images or both as 'perfect' or majority of the digital image analysts have to depend on the physical hardware available such as scanner and digitizer. In reality, it is not easy to obtain a perfect reference image because pre-processing procedures always resulted the existence of noise in the reference image. Wang & Bovik [11] has also pointed out that the assumption of perfect reference is reasonable only for image/video coding and communication applications.

In this paper, we propose a new statistical-based image quality metric (IQM) in which both reference image $X(i, j)$ and processed image $Y(i, j)$ are subjected to errors. As such the relationship between images X and Y may be modeled by the unreplicated linear functional relationship (ULFR) model [12], and it's coefficient of determination (COD) [13] is an indicator of performance. Definition and derivations of the ULFR model and it's COD are given in Section III. A simulation image is generated in Section IV to investigate the sensitivity of the proposed metric. Experiments using JPEG compression and de-noising filters are devoted in Section V

Manuscript received December 2008.

Chang Yun Fah, is with the Department of Mathematical and Acturial Sciences, Faculty of Information and Communication Technology, University of Tunku Abdul Rahman, 46200 Petaling Jaya, Selangor, MALAYSIA. (phone: 603-7955-1511; fax: 603-7955-1611; e-mail: changyf@mail.utar.edu.my).

Omar Mohd. bin Rijal, is with the Institute of Mathematical Sciences, University of Malaya, 50603 Kuala Lumpur, MALAYSIA. (phone: 603-7967-4323; fax: 603-7967-4143; e-mail: omarrija@um.edu.my).

Syed Abdul Rahman bin Syed Abu Bakar, is with the Faculty of Electrical Engineering, Universiti Teknologi Malaysia, 81310 Skudai, Johor, MALAYSIA. (phone: 607-553-5220; fax: 607-556-6272; e-mail: syed@fke.utm.my).

and conclusion is presented in Section VI. We will begin with some literature review on the existing IQMs in the Section II.

II. Literature Review

Over the years, there are a number of surveys and comparisons done on the image quality metrics (IQM). Michael et al. [14] discussed quality metrics applied to the still image compression problem. This paper classified the objective quality into four major categories, which are Mathematical metrics, metrics which incorporate the CSF and luminance adaptation, Metrics which incorporate observer preferences for suprathreshold artefact and Threshold perceptual metrics. Zhang [15] had discussed the assessment of image segmentation using empirical goodness methods and empirical discrepancy methods. Meanwhile, Eskiciogu and Fisher [16] had compared the performance of a set of metric under the compression problems. Avcibas et al. [17] considered the objective image quality metrics using Pixel different-based, non-statistical Correlation-based, Edge-based, Spectral distance-based, Context-based and HVS-based measures. On the other hand, Rubner et al. [18] divided the dissimilarity measures into 2 categories, which are the Distance Measure for Histograms and Non parametric test statistic. And, Eskiciogu [19] divided the subjective quality metrics into absolute and comparative categories, while the objective quality metrics are divided into numerical and graphical categories. Finally, Kim et al. [20], and Kinape and Amorim [21] discussed generally on a selected quality metrics.

The above studies showed that statistical-based IQMs have not been fully studied and their potential contributions have been overlooked. In fact, our reviewed on 278 related research papers reveal that pixel-based approach has the largest amount of image quality metrics for the past 26 years from year 1980 to 2006 (see Table 1). This is followed by the statistical-based and structural-based approaches.

IQM metrics	FR	RR	NR	Total
Statistical based	55	0	11	66
Pixel based	78	1	9	88
Structural based	40	1	1	42
Neighborhood based	14	0	1	15
HVS based	19	0	1	20
Graphical based	11	0	0	11
Subjective	24	0	0	24
Unclassified	11	0	1	12
Total	252	2	24	278

Table 1: Number of full reference image quality metrics from year 1980 to 2006 according to different approaches.

It was found that the number of statistical-based IQMs has increased over the mentioned periods as shown in Figure 1. There are a total of 55 FR and 11 NR statistical-based IQMs. In particular, there are seven statistical-based IQMs in years 1980 to 1989. The number of new statistical-based IQMs was increased to 26 from years 1990 to 1999 and it recorded 33 new statistical-based IQMs in the first six years of the twenty first century. We also noted that there is no study carried out on Statistical-based IQM for Reduced Reference (RR) until year 2006.

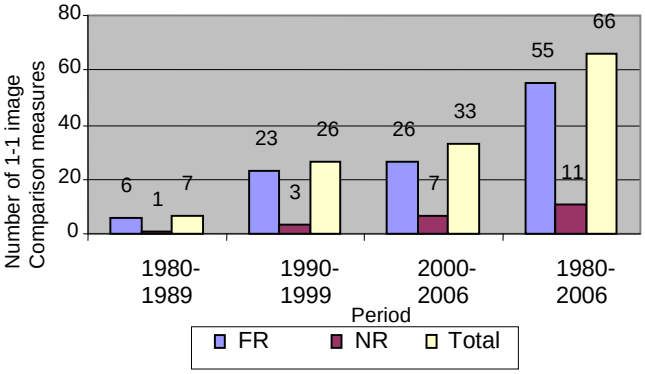

Fig 1: Summary of the FR and NR statistical-based IQM from year 1980 – 2006.

From our literature review, we discovered two problems common to the existing image quality metrics. They are

1. the existing full reference image quality metrics regarded one of the reference image and processed image, or both as perfect image. However, in reality, it is not easy to obtain a perfect reference image because pre-processing procedures always resulted in the existence of noise in the reference image.
2. some full reference image quality measures are good for certain type of images only, they will gave inconsistent interpretation of the quality index for different types of images.

The following sections propose a solution to accommodate the above-mentioned problems.

III. Unreplicated Linear Functional Relationship Model

A. ULFR Model

The conventional linear regression model assumes that the explanatory variable is fixed. However, this may not be realistic in many situations [22]. For the example in image processing, image manipulations occur in different preprocessing stages including image acquisition and we may think of $X(i, j)$ as having 'inherited' noise from the preprocessing stage. This suggests a new relationship between the two variables called ULFR model (see e.g. [12], [22], [23]).

Re-label the pixel values of the processed image $\{Y(i, j);\ \ i = 1, 2, \ldots, M,\ j = 1, 2, \ldots, N\}$ as $y_1, y_2, \ldots, y_{MN}$, the pixel values of the reference image

$\{X(i,j);\ \ i=1,2,\ldots,M,\ j=1,2,\ldots,N\}$ as $x_1,x_2,\ldots,x_{MN}$, and the true $Y(i,j)$ and $X(i,j)$ values will be denoted by $Y_1,Y_2,\ldots,Y_{MN}$ and $X_1,X_2,\ldots,X_{MN}$, respectively.

Suppose that now the *X* and *Y* are two linearly related unobservable variables

$$Y_i = \alpha_F + \beta_F X_i \tag{1}$$

and the two corresponding random variables *y* and *x* are observed with error ε and δ respectively

$$\left.\begin{aligned} y_i &= Y_i + \varepsilon_i \\ x_i &= X_i + \delta_i \end{aligned}\right\} \quad i=1,2,\ldots,MN \tag{2}$$

where α_F is the intercept and β_F is the slope of the functional model, and $\varepsilon_i \sim N(0,\sigma_\varepsilon^2)$ and $\delta_i \sim N(0,\sigma_\delta^2)$ are mutually independent and normally distributed random variables with mean 0 and constant, but with different variances. Model in Equations (1) and (2) is known as the unreplicated linear functional relationship (ULFR) when there is only one relationship between the two variables *Y* and *X*. This model yields the 2-vector $(x_i,y_i)'$, $i=1,2,\ldots,MN$ distributed as

$$\begin{pmatrix} x_i \\ y_i \end{pmatrix} \sim N\left[\begin{pmatrix} X_i \\ Y_i \end{pmatrix}, \begin{pmatrix} \sigma_\delta^2 & 0 \\ 0 & \sigma_\varepsilon^2 \end{pmatrix}\right] \tag{3}$$

and the likelihood function is given by

$$\begin{aligned} &L(\beta,\alpha,\sigma_\delta^2,\sigma_\varepsilon^2,X_1,\cdots,X_n) \\ &= \prod_{i=1}^{n} \frac{1}{2} \exp - \left[\frac{(\quad)^2}{2\sigma_\delta^2} + \frac{(y_i-\alpha-\beta X_i)^2}{2\sigma_\varepsilon^2} \right] \\ &= \frac{1}{(2\pi\,\sigma_\delta\sigma_\varepsilon)^n} \exp - \left[\frac{\sum(\quad)^2}{2\sigma_\delta^2} + \frac{\sum(\quad)^2}{2\sigma_\varepsilon^2} \right] \end{aligned} \tag{4}$$

If the ratio of the error variances $\frac{\sigma_\varepsilon^2}{\sigma_\delta^2} = \lambda$ is known and taking log for Equation (4), then the maximum likelihood estimators for α_F, β_F, σ_ε^2 and X_i are given as follows:

$$\hat{\alpha}_F = \bar{y} - \hat{\beta}_F \bar{x} \tag{5}$$

$$\hat{\beta}_F = \frac{\left(S_{yy} - \lambda S_{xx}\right) + \left\{\left(S_{yy} - \lambda S_{xx}\right)^2 + 4\lambda S_{xy}^2\right\}^{1/2}}{2S_{xy}} \tag{6}$$

$$\hat{\sigma}_\delta^2 = \frac{1}{n-2}\left[\lambda \sum \left(x_i - \hat{X}_i\right)^2 + \frac{1}{\lambda} \sum \left(y_i - \hat{\alpha}_F - \hat{\beta}_F \hat{X}_i\right)^2 \right] \tag{7}$$

and $$\hat{X}_i = \frac{\lambda x_i + \hat{\beta}_F\left(y_i - \hat{\alpha}\right)}{\lambda + \hat{\beta}_F^2} \tag{8}$$

where $\bar{x} = \frac{\sum x_i}{n}$, $\bar{y} = \frac{\sum y_i}{n}$, $S_{xx} = \sum (x_i - \bar{x})^2$, $S_{yy} = \sum (y_i - \bar{y})^2$ and $S_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y})$.

B. Functional Image Quality and Performance Metric

In order to measure the quality of JPEG compressed image and the performance of de-noising filters, we propose the COD for ULFR model as functional image quality metric. The construction of the COD for ULFR model has been developed in our previous work in Chang et al. [13].

The Equations (1) and (2) can be rewritten as $y_i = \alpha_F + \beta_F X_i + \varepsilon_i$ for $i=1,2,\ldots,n$.

If we substitute X_i by $(x_i - \delta_i)$, then we have the expression

$$\begin{aligned} y_i &= \alpha_F + \beta_F x_i + (\varepsilon_i - \beta_F \delta_i) \\ &= \alpha_F + \beta_F x_i + V_i \end{aligned} \tag{9}$$

where the errors of the model $V_i = (\varepsilon_i - \beta_F \delta_i) = y_i - (\alpha_F + \beta_F x_i)$, for $i=1,2,\ldots,n$ is a normally distributed random variable with zero mean and variance $\sigma_\varepsilon^2 + \beta_F^2 \sigma_\delta^2$. If $\hat{\alpha}_F$ and $\hat{\beta}_F$ are the estimates of α_F and β_F respectively, then

$$\hat{V}_i = y_i - \hat{y}_i = y_i - (\hat{\alpha}_F + \hat{\beta}_F x_i) \tag{10}$$

for $i=1,2,\ldots,n$, will be the residual of the model. From [23] and [24] that the sum of squared distances of the observed points from the fitted line or the residual sum of squares (SS_E) is given as:

$$\begin{aligned} \mathrm{SS}_E &= \frac{\sum \left\{ y_i - \left(\hat{\alpha} + \hat{\beta}_F x_i\right)\right\}^2}{\left(\lambda + \hat{\beta}_F^2\right)} \\ &= \frac{\mathrm{S}_{yy} - 2\hat{\beta}_F \mathrm{S}_{xy} + \hat{\beta}_F^2 \mathrm{S}_{xx}}{\left(\lambda + \hat{\beta}_F^2\right)} \end{aligned} \tag{11}$$

We shall consider here that the ratio of the error variances is equal to one ($\lambda = 1$). For those cases when $\lambda \neq 1$, we can always reduce this to the case of $\lambda = 1$ by dividing the observed values of y by $\lambda^{1/2}$ (see e.g. [23]). Hence, we have

$$\mathrm{SS}_E = \frac{\mathrm{S}_{yy} - 2\hat{\beta}_F \mathrm{S}_{xy} + \hat{\beta}_F^2 \mathrm{S}_{xx}}{1 + \hat{\beta}_F^2} \tag{12}$$

and the regression sum of squares which can be derived as

$$\begin{aligned} \mathrm{SS}_R &= \mathrm{S}_{yy} - \mathrm{SS}_E \\ &= \mathrm{S}_{yy} - \frac{\mathrm{S}_{yy} - 2\hat{\beta}_F \mathrm{S}_{xy} + \hat{\beta}_F^2 \mathrm{S}_{xx}}{1 + \hat{\beta}_F^2} \\ &= \frac{\hat{\beta}_F^2 \mathrm{S}_{yy} + 2\hat{\beta}_F \mathrm{S}_{xy} - \hat{\beta}_F^2 \mathrm{S}_{xx}}{1 + \hat{\beta}_F^2} \\ &= \frac{\hat{\beta}_F^2\left(\mathrm{S}_{yy} - \mathrm{S}_{xx}\right) + 2\hat{\beta}_F \mathrm{S}_{xy}}{1 + \hat{\beta}_F^2} \end{aligned} \tag{13}$$

In the same way as ordinary linear regression, we can now define the COD of the ULFR $\left(R_F^2\right)$ as the proportion of variation explained by the variable *x*, that is

$$\begin{aligned} R_F^2 &= \frac{SS_R}{S_{yy}} \\ &= \frac{S_{yy} - SS_E}{S_{yy}} \\ &= \frac{\mathrm{S}_{yy} - \dfrac{\mathrm{S}_{yy} - 2\hat{\beta}_F \mathrm{S}_{xy} + \hat{\beta}_F^2 \mathrm{S}_{xx}}{1+\hat{\beta}_F^2}}{S_{yy}} \\ &= \frac{\mathrm{S}_{yy} - \mathrm{S}_{yy} + 2\hat{\beta}_F \mathrm{S}_{xy} - \hat{\beta}_F^2 \mathrm{S}_{xx}}{S_{yy}\left(1+\hat{\beta}_F^2\right)} \\ &= \frac{\hat{\beta}_F\left(\hat{\beta}_F \mathrm{S}_{yy} - \hat{\beta}_F \mathrm{S}_{xx} + 2\mathrm{S}_{xy}\right)}{\mathrm{S}_{yy}\left(1+\hat{\beta}_F^2\right)} \\ &= \frac{\hat{\beta}_F\left(\hat{\beta}_F^2 \mathrm{S}_{xy} + \mathrm{S}_{xy}\right)}{\mathrm{S}_{yy}\left(1+\hat{\beta}_F^2\right)} \\ &= \frac{\hat{\beta}_F \mathrm{S}_{xy}\left(\hat{\beta}_F^2 + 1\right)}{\mathrm{S}_{yy}\left(1+\hat{\beta}_F^2\right)} \\ &= \frac{\hat{\beta}_F \mathrm{S}_{xy}}{\mathrm{S}_{yy}} \end{aligned} \tag{14}$$

C. Properties of the Functional Image Quality Metric

There are several properties of the proposed functional quality metric. These properties verify the criteria for a good quality metric as proposed by [6].

1) Reflexivity property

From the regression sum of squares, we obtained the range of the functional quality metric as $0 \le R_F^2 = \dfrac{\mathrm{SS}_R}{\mathrm{S}_{yy}} \le 1$. The R_F^2 has output one when the two images *X* and *Y* are identical and it approaches zero when the two images are dislike. This has been shown in the below experiments when the JPEG compression factor is at 100 and the noise level is at minimum.

2) Reaction to enlightening or darkening

Since image enlightening or darkening only involved constant translation of the intensity values, as other statistical correlation measures, the R_F^2 value, remain high for enlightens or darkens of an image. However, the R_F^2 value will drop with respect to an increasing enlightening or darkening percentage because more intensity values have converted to the value 1 (white color) or 0 (black color).

3) Reaction to binary images

When a IQM is applied to measure the quality of a binary image, the quality index shall returns a value between 0 and 1, but not only 0 or 1. This property is clearly shown in Section IV when a binary image was generated to examine the sensitivity of R_F^2. The results indicated that R_F^2 returns a value between 0 and 1.

4) Non-Symmetrical

To examine the symmetry properties of the R_F^2, we consider a new ULFR model by replacing Equation (1) with

$$X_i = \alpha_F^* + \beta_F^* Y_i \tag{15}$$

This could be happened when $S_{yy} > S_{xx}$. It can be shown that the estimated slope $\left(\hat{\beta}_F^*\right)$ and coefficient of determination, say $\tilde{R}_F^2$ for the new model when $\lambda = 1$ are

$$\hat{\beta}_F^* = \frac{\left(S_{xx} - S_{yy}\right) + \left\{\left(S_{xx} - S_{yy}\right)^2 + 4S_{xy}^2\right\}^{1/2}}{2S_{xy}} \tag{16}$$

and

$$\begin{aligned} \tilde{R}_F^2 &= \frac{\hat{\beta}_F^* S_{xy}}{S_{xx}} = \frac{\left(S_{xx} - S_{yy}\right) + \sqrt{\left(S_{xx} - S_{yy}\right)^2 + 4S_{xy}^2}}{2S_{xx}} \\ &= \frac{-\left(S_{yy} - S_{xx}\right) + \sqrt{\left(S_{yy} - S_{xx}\right)^2 + 4S_{xy}^2}}{2S_{xx}} \\ &= -\frac{1}{2S_{xx}}\left[\left(S_{yy} - S_{xx}\right) - \sqrt{\left(S_{yy} - S_{xx}\right)^2 + 4S_{xy}^2}\right] \\ &= -\frac{1}{2S_{xx}}\left[\left(S_{yy} - S_{xx}\right) - 2S_{yy}R_F^2 + \left(S_{yy} - S_{xx}\right)\right] \\ &= \frac{S_{yy}}{S_{xx}} R_F^2 - \frac{\left(S_{yy} - S_{xx}\right)}{S_{xx}} \end{aligned} \tag{17}$$

Let $S_{yy} = kS_{xx}$ and $k > 1$, then $\tilde{R}_F^2 = k\left(R_F^2 - 1\right) + 1$. (18)

Thus, our proposed R_F^2 is a non-symmetric metric. However, the failure of satisfying symmetric property does not create much problem to our method since S_{yy} and S_{xx} can always be determined and conversion between R_F^2 and $\tilde{R}_F^2$ can be easily done.

For consistency purpose, we consider the image with smaller variance as reference image and the image with larger variance as processed (compressed or de-noised) image.

IV. Sensitivity of the Quality and Performance Metric

It is important to realize that all existing IQMs only indicate the quality of a distorted image as compared with the reference image. In other words, these IQMs only carry the information about the similarity or distance between the two images. They do not reflect the amount of distortion in the distorted image. For example, if a quality measure shows the value of 0.5, it does not mean that 50% of the image is corrupted. Furthermore, we learnt that some IQMs provide

good indication for certain type of images, but they show inconsistency in interpreting for different types of images.

In this section we introduce a simple approach to investigate the sensitivity of the quality and performance metric based on the total distorted area. It does not only provides a consistent interpretation for different types of image, but it also give a physical meaning to our functional image quality and performance metric. A simple binary image J with size 100×100 is created, as shown in Figure 2. Let n_d be the number of distorted pixels randomly generated into the image J. For each given n_d, a set of 50 different distorted images is obtained and a mean performance index is calculated. Image *J1*, *J2* and *J3* in Figure 1 are examples of the distorted image for $n_d = 100,\ 2000$ and 10000, respectively. Figure 3 shows the relationship between the performance indices (or quality value) and the percentage of distorted areas, η for the proposed functional performance metric and two selected performance metrics; the current and well-established method, MSSIM and the classical method, RMSE. The fitted equations for these relationships are given as follow:

$$\mathrm{MSSIM}(\eta) \approx R_F^2(\eta) = \begin{cases} 1 & ,\eta \le 0.9607 \\ 1.0194\exp(-0.02\eta) & ,0.9607 < \eta \le 100 \end{cases} \tag{19}$$

and

$$\mathrm{RMSE}(\eta) = -0.00004\eta^2 + 0.0081\eta + 0.0064 \tag{20}$$

where $\eta = \dfrac{n_d}{MN}\times100\%$, and $M \times N$ is the size of the image.

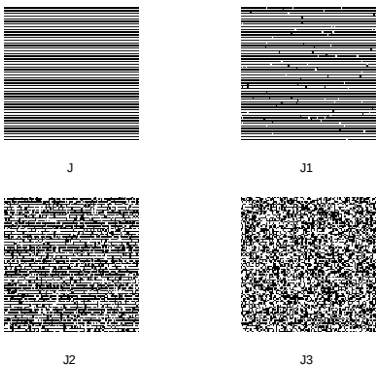

Fig. 2: Generated image and distorted images. J: simple binary image, J1: distorted image with n_d = 100, J2: distorted image with n_d = 2000, J3: distorted image with n_d = 10000.

It could be shown that the fitted equations for mean structural similarity index (MSSIM) and the proposed functional quality metric R_F^2 are very close to each others and they have an inverse exponential function to the percentage of distorted area. While, the mean square error (RMSE) is modeled by a quadratic equation. Similar estimated values could be obtained with different binary image sizes.

In general, the above result shows that the proposed metric is sensitive to the image distortions. With this model, we can find the desired index value by setting the percentage of distorted area in controlling the quality of image or we can estimate the percentage of distorted area from the performance index obtained using the reverse of Equation (19) as follow.

$$\eta = \begin{cases} 100 & ,R_F^2 \le 0.1379608 \\ -50\ln\left(\dfrac{R_F^2}{1.0194}\right) & ,0.1379608 < R_F^2 \le 1 \end{cases} \tag{21}$$

For example, if we try to keep the quality of an image to be less than 10% of the total number of distorted areas (or $n_d = 1000$ distorted pixels in this case) from the original image, then the minimum performance index for R_F^2 value is 0.8187 calculated from Equation (19).

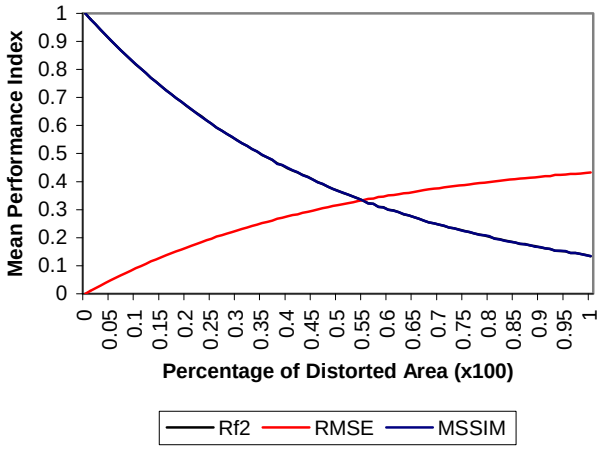

Fig. 3: Relationship between the mean quality value (or average performance index) and the percentage of distorted area. Note that Rf2= R_F^2.

V. Experiments and Results

Two experiments are discussed. In the first experiment we will investigate and compare different metrics as performance indicator for JPEG compression. The proposed metric is compared with some selected objective metrics based on their popularity. The second experiment studies the performance of some noise removal techniques to remove Gaussian noise. The proposed metric is used to compare the quality of enhanced image obtained from the noise removal techniques with the original image. A set of six frequently used standard test images from the literature review is being considered in these studies. These standard test images are Lena, Baboon, Airplane, Bridge, Boat and Peppers (see Figure 4). These test images are saved in bitmap format and converted to grey scale

image if they are colored.

Fig. 4: Standard test images. Top (left to right): Lena (size 256×256), Baboon (size 256×256), Airplane (size 512×512). Bottom (left to right): Bridge (size 145×145), Boat (size 512×512), Peppers (size 512×512).

A. Image Quality Assessment for JPEG Compression

Figures 5 to 10 show the plots of performance (quality) index versus the JPEG compression factor of range 1 to 100 obtained from the test images for various performance metrics, R_F^2, R_S^2, MSSIM and RMSE. It is expected that higher compression factor produces greater compressed image quality (better performance) with higher similarity value or lower distance value. The performance indices are nearly 1 (and zero for RMSE) when the compression factor is 100. The performance indices decreased with respect to a decreasing (increasing for RMSE) compression factor.

It is observed that the performance metrics work reasonably well except for the RMSE. As the JPEG compression factor increase (or the percentage of discard decrease), the performance metrics R_F^2, R_S^2 and MSSIM show an increasing index in compressed quality. Like the peak signal to noise ratio and mean square error methods, the RMSE is not correlated well with human visual system. This is reflected from the RMSE plots that the distances between decompressed image and original image remain very low as the compression factor decreased.

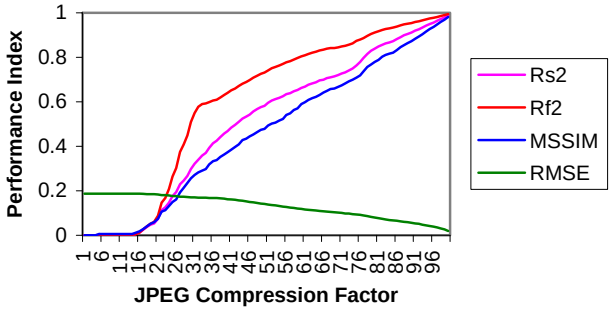

Fig. 5: Plot of performance index versus compression factor for Lena test image. Note that Rf2= R_F^2 and Rs2= R_S^2.

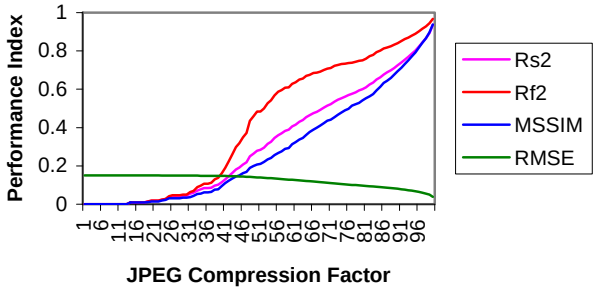

Fig. 6: Plot of performance index versus compression factor for Baboon test image. Note that Rf2= R_F^2 and Rs2= R_S^2.

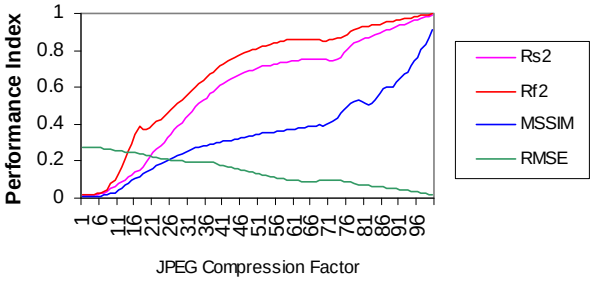

Fig. 7: Plot of performance index versus compression factor for Airplane test image. Note that Rf2= R_F^2 and Rs2= R_S^2.

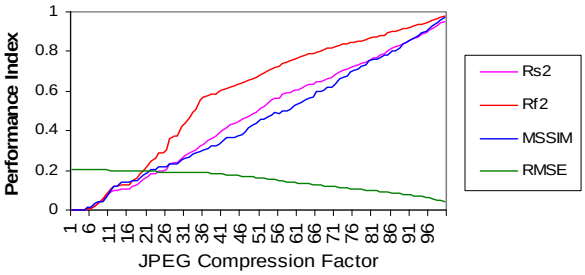

Fig. 8: Plot of performance index versus compression factor for Bridge test image. Note that Rf2= R_F^2 and Rs2= R_S^2.

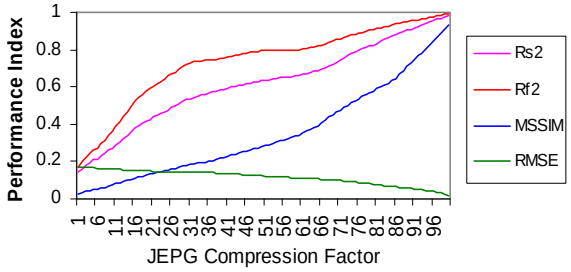

Fig. 9: Plot of performance index versus compression factor for Boat test image. Note that Rf2= R_F^2 and Rs2= R_S^2.

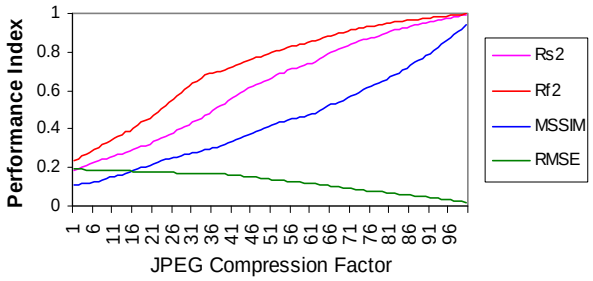

Fig. 10: Plot of performance index versus compression factor for Peppers test image. Note that Rf2= R_F^2 and Rs2= R_S^2.

Figures 5 to 10 indicated that R_F^2, R_S^2 and MSSIM are good performance indicators for JPEG compression. Unfortunately, these plots have no practical application due to two problems: (i) JPEG compression factor is usually unknown, and (ii) The same JPEG compression factor may be resulted in different performance index for different types of images. Figure 11 compares the original test images with their decompressed images at compression factors 74 and 50, respectively. Obviously, for instance, the decompressed Lena image has better visual quality as compared to Baboon image at a given compression factor. Instead, Lena decompressed image at compression factor 50 $\left(R_F^2 = 0.7320\right)$ has 'similar' image quality as Baboon decompressed image at compression factor 74 $\left(R_F^2 = 0.7316\right)$.

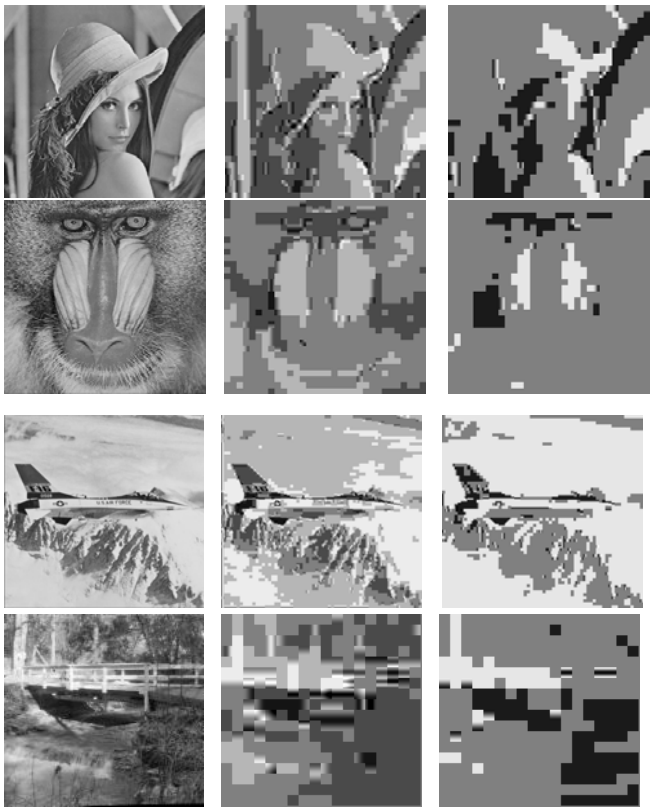

Fig. 11: Lena, Baboon, Airplane, Bridge, Boat and Peppers (from top to bottom) images: original, decompressed image with factor 74, and decompressed image with factor 50 (From left to right).

To overcome these problems, Equation (21) may be used to provide a consistent interpretation for the distortion area. We now estimate the percentage of distorted areas produced by JPEG compression for Lena test image. The experiment shows that the R_F^2 values obtained from the JPEG compression factors 90, 80, 50 and 25 are 0.9543, 0.9112, 0.7320 and 0.2087, respectively. By substituting these R_F^2 values into Equation (21), we found that the above mentioned compression factors produced 3.30%, 5.61%, 16.56% and 79.30% of distortion areas, respectively. Similarly, the percentage of distorted areas produced by JPEG compression for Baboon test image at compression factors 90, 80, 50 and 25 are 9.60%, 15.21%, 37.34% and 100%, respectively. Figure 12 to Figure 17 show the plots of the percentage of distorted area versus R_F^2 values for all test images. Given a R_F^2 value, these plots show a consistent indication for the amount of distortion generated by JPEG compression, regardless of the image tested. For instance, when R_F^2 value is about 0.834, JPEG compression generated about 10% distorted area (pixels) on all tested images, which the results are corresponding to their image size.

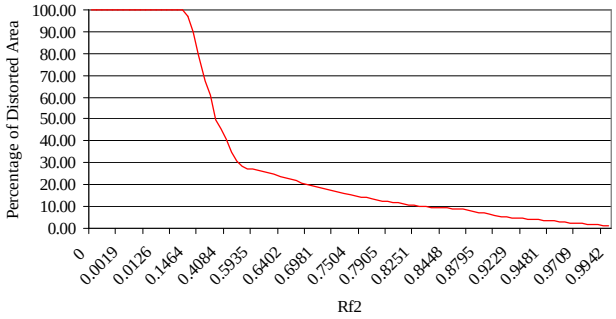

Fig. 12: Plot of the percentage of distorted area versus R_F^2 value for Lena image. Note that Rf2= R_F^2.

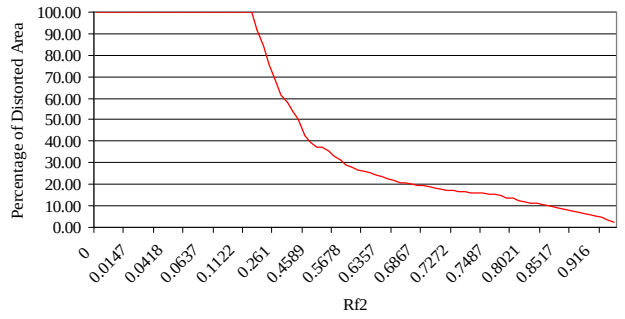

Fig. 13: Plot of the percentage of distorted area versus R_F^2 value for Baboon image. Note that Rf2= R_F^2 .

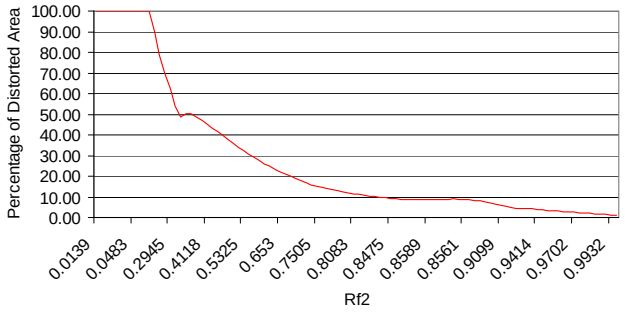

Fig. 14: Plot of the percentage of distorted area versus R_F^2 value for Airplane image. Note that Rf2= R_F^2 .

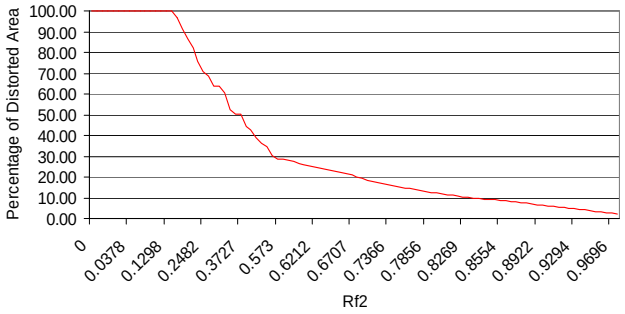

Fig. 15: Plot of the percentage of distorted area versus R_F^2 value for Bridge image. Note that Rf2= R_F^2 .

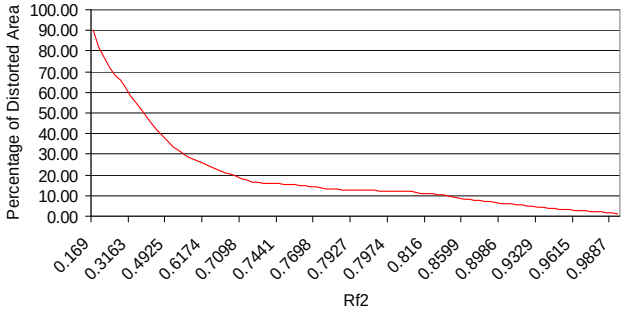

Fig. 16: Plot of the percentage of distorted area versus R_F^2 value for Boat image. Note that Rf2= R_F^2 .

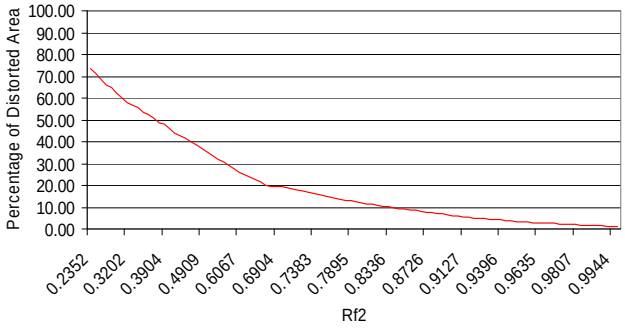

Fig. 17: Plot of the percentage of distorted area versus R_F^2 value for Peppers image. Note that Rf2= R_F^2 .

B. Performance Evaluation for Image De-noising Filters

The second experiment used R_F^2 to evaluate the performance of some de-noising filters. Gaussian noise with mean zero and increasing variance values from 0 to 0.029 were added into the original image. De-noising filters are then applied to enhance the quality of the distorted image. Ten de-noising filters are tested in these experiments. Five of them are low-pass filter such as Average filter, Circular averaging filter, Gaussian low-pass filter, Motion filter and Median filter. Others are high-pass filters, i.e. Laplacian of Gaussian filter, approximating the 2-D Laplacian operator, Prewitt horizontal edge-emphasizing filter, Sobel horizontal edge-emphasizing filter and Unsharp contrast enhancement filter. It is well known that high-pass filters are unable to remove Gaussian noise and Salt & Pepper noise. The high-pass filters are introduced here to proof that the proposed metric is sensitive in differentiating the performance of low-pass and high-pass filters. These filters are obtained from MATLAB 7.1 with a default parameters applied. For each filtered image, its similarity value compare with the reference image (original image before add noise) is calculated from R_F^2 .

Figure 18 and Figure 19 show the plot of R_F^2 values for the ten de-noising filters of Lena and Baboon test images. The R_F^2 values clearly show that all the high-pass filters produced bad-quality enhanced image with low similarity value. The experiment shows that the low-pass filters produced much better result, in which their enhanced image quality is closer to the original image. Note that the performance of Average filter is slightly better than other low-pass filters when the image was seriously distorted by Gaussian noise and it is consistent with all levels of Gaussian noise. The quality of enhanced image produced by Circular averaging filter and Motion filter are both lower than Average filter. While, the performance of Gaussian low-pass filter decreased as the variance of Gaussian noise increased, although it produced the highest enhanced image quality at lower level of Gaussian noise. Generally, the Average filter outperforms other low-pass filters and it is considered the most ideal filter for removing Gaussian noise. The Average filter produced an

enhanced image with R_F^2 value of at least 0.9387 or less than 4.12% distorted area calculated from Equation (21).

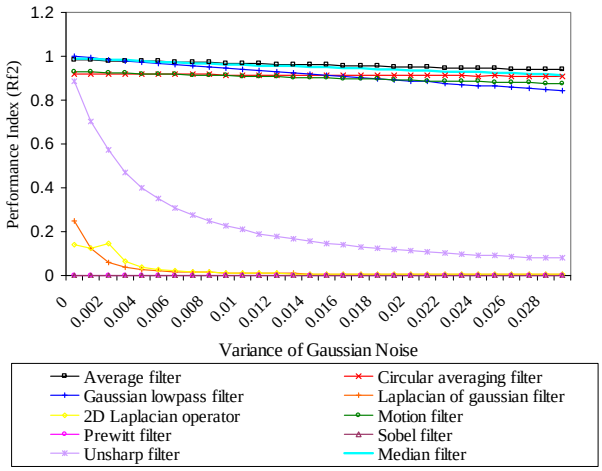

Fig. 18: Plot of R_F^2 (=Rf2) values against variances of Gaussian noise for different de-noising filters from Lena test image.

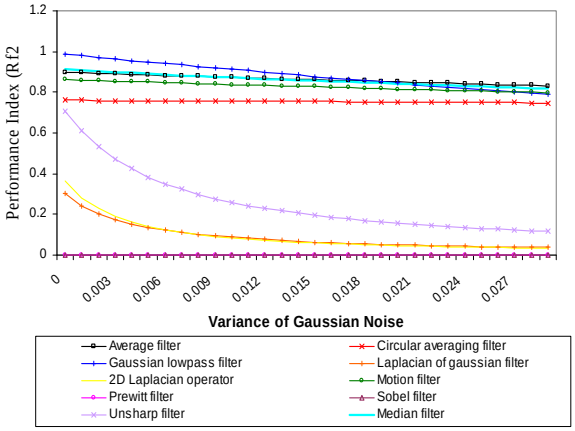

Fig. 19: Plot of R_F^2 (=Rf2) values against variances of Gaussian noise for different de-noising filters from Baboon test image.

A similar experiment to remove Salt & Pepper noise was conducted and its results are depicted in Figure 20 and Figure 21 for Lena and Baboon test images, respectively. Again, we notice that the high-pass filters perform badly in removing Salt & Pepper noise. Obviously, the low-pass Median filter is the best and consistent technique to remove Salt & Pepper noise at all levels of noise density. At noise density 0.3, for example, the enhanced image obtained from Median filter still remain at high quality as compared to the original image with $R_F^2 = 0.9223$ or about 5% distorted area calculated from Equation (21).

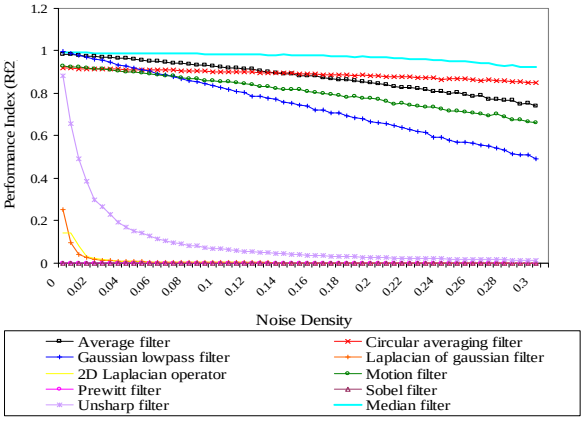

Fig. 20: Plot of R_F^2 (=Rf2) values against density of Salt & Pepper noise for different de-noising filters from Lena test image.

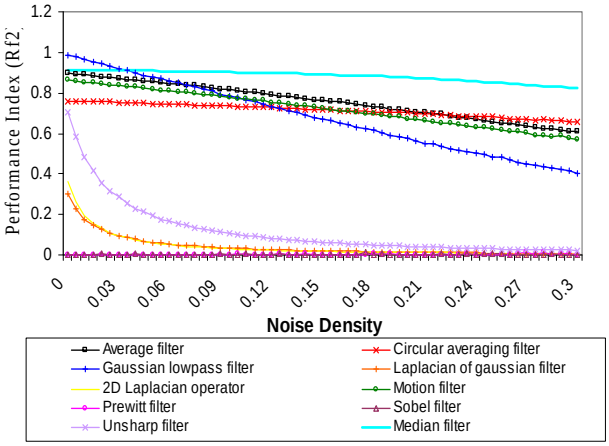

Fig. 21: Plot of R_F^2 (=Rf2) values against density of Salt & Pepper noise for different de-noising filters from Baboon test image.

VI. Conclusions

As a conclusion from this study, the proposed R_F^2 is used as the performance indicator for JPEG compression and de-noising filters. In JPEG compression, we observed that the R_F^2 perform as good as R_S^2 and MSSIM. However, we believe that the proposed metric is more realistic as it considered both the reference image and processed image are random variables. The MSSIM also performed well in these examples, but a predefined parameter setting [5] from the user is needed for different applications and images. This is not possible when the properties of the images are unknown or the performance metric is implemented to an automatic system.

Besides, this study provides an easy and consistent way of interpreting the R_F^2 value in the context where the percentage of distorted area is generated by JPEG compression. With the given formula, it is now possible to

estimate the amount of distortion in an image from its R_F^2 value, regardless of the image tested.

Lastly, the proposed R_F^2 is applied to identify the suitable filter for removing Gaussian noise and Salt and Pepper noise. The results indicate that average filter produced the most desired and consistent results than other selected type of filters in removing the Gaussian noise. In the case of Salt and Pepper noise, Median filter produces the highest quality of filtered image on almost all levels of noise density.

REFERENCES

[1] Z. Wang and A.C. Bovik, "A universal image quality index," *Signal Processing Letter*, vol. 9, no. 3, 2002, pp. 81 – 84.

[2] R.C. Gonzalez and R. E. Woods, *Digital Image Processing*, New York: Addison Wesley, 1992, pp. 165 – 166.

[3] J.G. Puttenstein, I. Heynderickx and G. de Haan, "Evaluation of objective quality measures for noise reduction in TV-systems," *Signal Processing: Image Communication*, vol. 19, 2004, pp. 109 – 119.

[4] Rogerio M. Kinape and Mardson F. Amorim, "A study of the most important image quality measures," in *Proceedings of 25th Annual International Conference of the IEEE EMBS*, 2003, pp. 934 – 936.

[5] 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, 2004, pp. 600 – 612.

[6] D. Van der Weken, M. Nachtegael and E.E. Kerre, "Image quality evaluation," in *Proceedings of 6th International Conference on Signal Processing*, vol. 1, Beijing, 2002, pp. 711 – 714.

[7] B.D. Steinberg, "A theory of the effect of hard limiting and other distortion upon the quality of microwave images," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 35, 1987, pp. 1462 – 1472.

[8] Ioannis K. Vlachos, George D. Sergiadis, "Image quality indices based on fuzzy discrimination information measures," *WSEAS Transactions on Circuits and Systems*, vol. 3, no. 2, April 2004, pp. 241-247.

[9] P. Planinsic, "Subband Coded images quality control using fuzzy logic and neural network," in *Proceedings of the WSEAS Internatinal Conference on Neural Network and Applications*, Tenerife, Spain, 2001, pp. 3051 – 3056.

[10] B.N. Kayani, A.M. Mirza and H. Iftikhar, "Pixel and feature level multi-resolution image fusion based on fuzzy logic," in *Proceedings of the 6th WSEAS International Conference on Wavelet Analysis and Multirate Systems*, Bucharest, Romania, 2006, pp. 88 – 91.

[11] Z. Wang, A.C. Bovik and L. Lu, "Why is image quality assessment so difficult?," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 4, Orlando, 2002, pp. 3313 – 3316.

[12] A.G. Hussin, "Pseudo-Replication in Functional Relationships With Environmental Applications," Ph.D. dissertation, University of Sheffield, 1997.

[13] Y.F. Chang, A.G. Hussin and O.M. Rijal, "An investigation of causation: the unreplicated linear functional relationship model," *ANSI Journal of Applied Sciences*, vol. 7, no. 1, 2007, pp. 20 – 26.

[14] P. Michael, Eckert and A.P. Bradley, "Perceptual quality metrics applied to still image compression," *Signal Processing*, vol. 70, no. 3, 1998, pp. 177 – 200.

[15] Y.J. Zhang, "A survey on evaluation methods for image segmentation," *Pattern Recognition*, vol. 29, no. 8, 1996, pp. 1335 – 1346.

[16] A.M. Eskiciogu and P.S. Fisher, "Image quality measures and their performance," *IEEE Transaction on Communications*, vol. 43, no. 12, 1995, pp. 2959 – 2965.

[17] I. Avcibas, B. Sankur and K. Sayood, "Statistical evaluation of image quality measures," *Journal of Electronic Imaging*, vol. 11, no. 2, 2002, pp. 206 – 223.

[18] Y. Rubner, J. Puzicha, C. Tomasi and J. M. Buhmann, "Empirical evaluation of dissimilarity measures for color and texture," *Journal of Computer Vision and Image Understanding*, vol. 84, no. 1, 2001, pp. 25 – 43.

[19] A.M. Eskicioglu, "Quality measurement for monochrome compressed images in the past 25 years," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 4, Istanbul, Turkey, 2000, pp. 1907 – 1910.

[20] J.S. Kim, M.S. Cho and B.T. Choi, "Study on the methods of digital image quality evaluation," in *IEEE Region 10 Conference* (TENCON), vol. 1, 2004, pp. 359 – 362.

[21] R.M. Kinape and M.F. Amorim, "A study of the most important image quality measures," in *Proceedings of 25th Annual International Conference of the Engineering in Medicine and Biology Society*, vol. 1, 2003, pp. 934 – 936.

[22] W.A. Fuller, *Measurement Error Models*, New York: John Wiley. 1987.

[23] M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics*, vol. 2, London: Griffin. 1961, pp. 399 – 440.

[24] T.W. Anderson, "The 1992 Wald memorial lectures: estimating linear statistical relationships," *Annals of Statistics*, vol. 12, 1984, pp. 1 – 18.