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 versa. 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^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.
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.

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.

<table>
<thead>
<tr>
<th>IQM metrics</th>
<th>FR</th>
<th>RR</th>
<th>NR</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistical based</td>
<td>55</td>
<td>0</td>
<td>11</td>
<td>66</td>
</tr>
<tr>
<td>Pixel based</td>
<td>78</td>
<td>1</td>
<td>9</td>
<td>88</td>
</tr>
<tr>
<td>Structural based</td>
<td>40</td>
<td>1</td>
<td>1</td>
<td>42</td>
</tr>
<tr>
<td>Neighborhood based</td>
<td>14</td>
<td>0</td>
<td>1</td>
<td>15</td>
</tr>
<tr>
<td>HVS based</td>
<td>19</td>
<td>0</td>
<td>1</td>
<td>20</td>
</tr>
<tr>
<td>Graphical based</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>Subjective</td>
<td>24</td>
<td>0</td>
<td>0</td>
<td>24</td>
</tr>
<tr>
<td>Unclassified</td>
<td>11</td>
<td>0</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>Total</td>
<td>252</td>
<td>2</td>
<td>24</td>
<td>278</td>
</tr>
</tbody>
</table>

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

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 \(x_i, x_{i2}, \ldots, x_{iN}\), and
the true \(Y(i, j)\) and \(X(i, j)\) values will be denoted by
\(Y_i, Y_{i2}, \ldots, Y_{iN}\) and \(X_i, X_{i2}, \ldots, X_{iN}\), respectively.

Suppose that now the \(X\) and \(Y\) are two linearly related
unobservable variables
\[Y_i = \alpha + \beta X_i\]  \(^{(1)}\)
and the two corresponding random variables \(y\) and \(x\) are
observed with error \(\varepsilon\) and \(\delta\) respectively
\[y_i = Y_i + \varepsilon_i, \quad x_i = X_i + \delta_i\]  \(^{(2)}\)
where \(\alpha\) is the intercept and \(\beta\) is the slope of the functional
model, and \(\varepsilon_i \sim N\left(0, \sigma^2\right)\) and \(\delta_i \sim N\left(0, \sigma^2\right)\) 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 \((x_i, y_i)^\prime\), \(i = 1, 2, \ldots, MN\) distributed as
\[
\begin{pmatrix} x_i \\ y_i \end{pmatrix} \sim N\left(\begin{pmatrix} X_i \\ Y_i \end{pmatrix}, \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}\right) \]
\(^{(3)}\)
and the likelihood function is given by
\[
L(\beta, \alpha, \sigma^2, \sigma^2, X_1, \ldots, X_n) = \prod_{i=1}^{N} \frac{1}{2\pi \sigma_\varepsilon \sigma_\delta} \exp\left[-\frac{1}{2\sigma^2_\varepsilon} \left(y_i - \alpha - \beta X_i\right)^2 - \frac{1}{2\sigma^2_\delta} \left(x_i - \alpha - \beta X_i\right)^2\right] \]
\(^{(4)}\)
If the ratio of the error variances \(\frac{\sigma^2_\varepsilon}{\sigma^2_\delta} = \lambda\) is known and taking
log for Equation (4), then the maximum likelihood estimators
for \(\alpha, \beta, \sigma^2\) and \(X\), are given as follows:
\[
\hat{\alpha} = \bar{y} - \hat{\beta} \bar{x} \]
\(^{(5)}\)
\[\hat{\beta} = \frac{\left(S_n - \lambda S_m\right) + \left(\left(S_n - \lambda S_m\right)^2 + 4\lambda S_m\right)^{\frac{1}{2}}}{2S_m} \]
\(^{(6)}\)
\[
\hat{\sigma}^2 = \frac{1}{n-2} \left[\lambda \sum (x_i - \bar{x})^2 + \lambda \sum (y_i - \hat{\alpha} - \hat{\beta} \bar{x})^2\right] \]
\(^{(7)}\)
and \(\hat{X}_i = \frac{\hat{\beta} x_i + (y_i - \hat{\alpha})}{\lambda + \hat{\beta}^2} \)
\(^{(8)}\)
where \(\bar{x} = \frac{\sum x_i}{n}, \quad \bar{y} = \frac{\sum y_i}{n}, \quad S_n = \sum (x_i - \bar{x})^2, \quad S_m = \sum (y_i - \bar{y})^2, \quad S_{\bar{y}} = \sum (y_i - \bar{y})^2 and S_{\bar{x}} = \sum (x_i - \bar{x})(y_i - \bar{y}).\)

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
\[y_i = \alpha + \beta x_i + \varepsilon_i, \quad \text{for } i = 1, 2, \ldots, n.\]
If we substitute \(X_i\) by \((x_i - \delta_i)\), then we have the expression
\[y_i = \alpha + \beta (x_i - \delta_i) + \varepsilon_i = \alpha + \beta x_i + \varepsilon_i\]
\(^{(9)}\)
where the errors of the model \(V_i = (\varepsilon_i - \beta x_i)\) for \(i = 1, 2, \ldots, n\) is a
normally distributed random variable with zero mean and
variance \(\sigma^2 + \beta^2 \sigma^2\). If \(\hat{\alpha}\) and \(\hat{\beta}\) are the estimates of \(\alpha\) and
\(\beta\) respectively, then
\[\hat{V}_i = y_i - \hat{y}_i = y_i - (\hat{\alpha} + \hat{\beta} x_i)\]
\(^{(10)}\)
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 (SSR)
is given as:
\[SSR = \sum \left(y_i - (\hat{\alpha} + \hat{\beta} x_i)^2 \right)^{\frac{1}{2}} \]
\[\frac{\lambda + \hat{\beta}^2}{\hat{\beta}^2} \]
\(^{(11)}\)
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 \(\lambda^{\frac{1}{2}}\) (see e.g. [23]). Hence, we have
\[SSR = \frac{S_m - 2\hat{\beta} S_m + \hat{\beta}^2 S_m}{1 + \hat{\beta}^2} \]
\(^{(12)}\)
and the regression sum of squares which can be derived as
\[SSR = SSR - SSR \]
\[SSR = \frac{S_m - 2\hat{\beta} S_m + \hat{\beta}^2 S_m}{1 + \hat{\beta}^2} \]
\[\frac{\hat{\beta}^2 (S_m - S_m) + 2\hat{\beta} S_m}{1 + \hat{\beta}^2} \]
\[^{(13)}\]
In the same way as ordinary linear regression, we can now define the COD of the ULFR \( R_F^2 \) as the proportion of variation explained by the variable \( x \), that is
\[
R_F^2 = \frac{SS_y}{SS_y} = \frac{S_{yy} - SS_y}{S_{yy}} = \frac{S_{yy} - S_{yy} - 2\hat{\beta}_yS_{xy} + \hat{\beta}_y^2S_{xx}}{S_{yy}} = \frac{S_{yy} - S_{yy} + 2\hat{\beta}_yS_{xy} - \hat{\beta}_y^2S_{xx}}{S_{yy} + \hat{\beta}_y^2S_{xx}} = \frac{\hat{\beta}_yS_{xy} + S_{xy}}{S_{yy} + \hat{\beta}_y^2S_{xx} - \hat{\beta}_yS_{xy}} = \frac{\hat{\beta}_yS_{xy}}{S_{yy}} \tag{14}
\]

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 \( 0 \leq R_F^2 = \frac{SS_y}{SS_y} \leq 1 \). 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 \( X_i = \alpha_x + \beta_x Y_i \). This could be happened when \( S_{yp} > S_{xp} \). It can be shown that the estimated slope \( \hat{\beta}_y \) and coefficient of determination, say \( \hat{R}_F^2 \) for the new model when \( \lambda = 1 \) are
\[
\hat{\beta}_y = \frac{S_{xy} - S_{xp}}{S_{xp}} + \sqrt{(S_{xy} - S_{xp})^2 + 4S^2_{xp}} \tag{15}
\]
and
\[
\hat{R}_F^2 = \frac{S_{xp}R_F^2 - (S_{xp} - S_{xp})}{S_{xp}} \tag{16}
\]

Let \( S_{xp} = kS_{yp} \) and \( k > 1 \), then \( \hat{R}_F^2 = k(R_F^2 - 1) + 1 \). \tag{17}

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 \( S_{yp} \) and \( S_{xp} \) can always be determined and conversion between \( R_F^2 \) and \( \hat{R}_F^2 \) 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 \( 100 \times 100 \) 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 \( J_1, J_2 \) and \( J_3 \) 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, \( \eta \) 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:

**MSSIM**

\[
MSSIM(\eta) = R^2_F(\eta) = \begin{cases} 
1 & , \eta \leq 0.9607 \\
1.0194 \exp(-0.02\eta) & , 0.9607 < \eta \leq 100 
\end{cases}
\]  

(19)

and

**RMSE**

\[
RMSE(\eta) = -0.00004\eta^2 + 0.0081\ln\eta + 0.0064
\]

(20)

where \( \eta = \frac{n_d}{MN} \times 100\% \), and \( M \times N \) is the size of the image.

\[
\eta = \frac{100}{1.0194} & , R^2_F \leq 0.1379608 \\
-50\ln\left(\frac{R^2_F}{1.0194}\right) & , 0.1379608 < R^2_F \leq 1
\]

(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 \( R^2_F \) value is 0.8187 calculated from Equation (19).

![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 \( n_d = 10000 \).](image)

It could be shown that the fitted equations for mean structural similarity index (MSSIM) and the proposed functional quality metric \( R^2_F \) 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.

**Percentage of Distorted Area (x100)**

- **Mean Performance Index**

![Fig. 3: Relationship between the mean quality value (or average performance index) and the percentage of distorted area. Note that R2F= R^2_F.](image)

**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 $R_f^2=R_f^2$ and $R_s^2=R_s^2$.

Fig. 6: Plot of performance index versus compression factor for Baboon test image. Note that $R_f^2=R_f^2$ and $R_s^2=R_s^2$.

Fig. 7: Plot of performance index versus compression factor for Airplane test image. Note that $R_f^2=R_f^2$ and $R_s^2=R_s^2$.

Fig. 8: Plot of performance index versus compression factor for Bridge test image. Note that $R_f^2=R_f^2$ and $R_s^2=R_s^2$.

Fig. 9: Plot of performance index versus compression factor for Boat test image. Note that $R_f^2=R_f^2$ and $R_s^2=R_s^2$. 
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 ($R_f^2 = 0.7320$) has ‘similar’ image quality as Baboon decompressed image at compression factor 74 ($R_f^2 = 0.7316$).

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. 10: Plot of performance index versus compression factor for Peppers test image. Note that $R_f^2 = R_f^2$ and $R_s^2 = R_s^2$.

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).

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

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

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).
values clearly show that all the high-pass filters produced a value for Airplane image. Note that $R_f^2 = R_f^2$. 

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

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

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

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

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

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 $R_f^2$ 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 $R_E^2$ value of at least 0.9387 or less than 4.12% distorted area calculated from Equation (21).

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

Fig. 19: Plot of $R_E^2$ (=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 $R_E^2 = 0.9223$ or about 5% distorted area calculated from Equation (21).

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

Fig. 21: Plot of $R_E^2$ (=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 $R_E^2$ is used as the performance indicator for JPEG compression and de-noising filters. In JPEG compression, we observed that the $R_E^2$ perform as good as $S_R^2$ 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 $R_E^2$ 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 $R_e^2$ value, regardless of the image tested.

Lastly, the proposed $R_e^2$ 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