Paper Title: A New Image Restoration Performance Measure with High Precision (Revised Manuscript)

Authors: Steven S. O. CHOY, Yuk-Hee CHAN and Wan-Chi SIU

Affiliation: Department of Electronic Engineering,
The Hong Kong Polytechnic University

Correspondence Address:
Dr. Yuk-Hee CHAN
Dept. of Electronic Engineering,
The Hong Kong Polytechnic University,
Hung Hom, Kowloon,
Hong Kong.

Phone No.: (852) 2766 6250 (Steven S. O. Choy)
(852) 2766 6224 (Yuk-Hee Chan)
(852) 2766 6229 (Wan-Chi Siu)

Fax No.: (852) 2362 8439

Email: ensochoy@hkpu02.polyu.edu.hk (Steven S. O. Choy)
enyhchan@hkpucc.polyu.edu.hk (Yuk-Hee Chan)
enwcsiu@hkpucc.polyu.edu.hk (Wan-Chi Siu)

Abstract

The field of image restoration lacks promising comparison vehicle for judging the effectiveness of competing algorithms. By far the most widely adopted quantitative measurement of image restoration performance is by means of the SNR improvement. In this paper we address the issues on performance assessment of image restoration and propose a unified framework for the performance measure. The SNR improvement, which is within the proposed framework, is shown to be an inappropriate performance measure for image restoration. By introducing a metric for pixel fidelity improvement and incorporating main properties of the human visual system into the measurement, we devise a performance measure of better quality, particularly of higher precision.

Subject terms: image restoration; restoration performance measure; precision evaluation.

A New Image Restoration Performance Measure with High Precision

1 Introduction

Image restoration refers to the problem of reconstructing or estimating an original image from its distorted rendition. It is an active area of research and finds its applications in many fields such as medical imaging, space imagery, forensic science and commercial imaging. While many new restoration algorithms, ranging from deterministic iterative methods to optimal stochastic filtering, have been proposed in the last decade [1], there is still no promising and effective comparison vehicle to judge the performance of those competing algorithms. It is critical that the area of image restoration lacks performance indices to indicate how much the improvement made by the newer and sophisticated algorithms [2].

By far the most popular quantitative measure of image quality is SNR (Signal-to-Noise Ratio). Therefore, *SNR improvement*, which is defined as the difference between the SNR of the restored image and the SNR of the distorted image, is widely adopted as a restoration performance index. Mathematically,

$$\text{SNR Improvement} = 10 \log \frac{\sum_{i,j}(x_{i,j} - y_{i,j})^2}{\sum_{k,l}(x_{k,l} - \hat{x}_{k,l})^2}, \tag{1}$$

where x, y and $\hat{x}$ are the original image, the distorted image and the restored image, respectively. This objective measure is usually applied to evaluate restoration performance, and have become a *de facto* standard in the comparative study of restoration algorithms [2-4]. However, it is well known that such a measure, which is based on the MSE criterion, does not agree with the properties of the human visual system. This implies that performance assessment by means of the SNR improvement is, to a certain extent, of low accuracy with respect to the human visual system. Besides, it is found by us that the SNR improvement is of low precision in evaluating restoration performance. Apart from accuracy and precision, the SNR improvement as a measurement of restoration performance conveys unclear message.

The objective of this paper is to devise a quantitative measure of image restoration performance. Our primary concern is the precision of the measure. It should be complex enough to describe restoration performance, but at the same time simple enough to allow efficient implementation of the measurement. The measure we propose comes from the idea that a weighted sum of *pixel fidelity improvement* can serve as a restoration performance index. In fact, it can be shown that SNR improvement is basically within the same framework with its own definition of pixel fidelity improvement and particular weighting coefficients. We are therefore convinced that, by finding a better metric for pixel fidelity improvement and incorporating main properties of the human visual system into the weighting coefficients, an improved performance measure that overcomes the main limitations of the SNR improvement can be achieved. This paper is arranged into several sections. In Section 2, the properties that a good performance measure should have, as well as the limitations of the SNR improvement, will be discussed in detail. The improved measure is then proposed, and the derivation is presented in Section 3. In Section 4, the proposed measure is evaluated and compared with the SNR improvement. Finally, conclusions are given in Section 5.

2 Measure of Image Restoration Performance

2.1 Formulation

Image restoration is a process to improve the quality of processed image. The performance of a restoration can thus be evaluated by measuring the amount of improvement in image quality. To measure that improvement, we need to have the original, the distorted and the restored images available in the measuring process. Measurement of restoration performance can generally be viewed as a process that intakes three image vectors, x (original), y (distorted) and $\hat{x}$ (restored), and returns a scalar value. This output value is thus a measurement used to indicate how much the image quality is improved from y to $\hat{x}$ with respect to x. In applications of image restoration, image quality usually refers to the image's fidelity to its original. Suppose the fidelity improvement of each individual pixel is quantified by certain means and denoted as $f_{i,j}$, where (i, j) represents the pixel location. Clearly, $f_{i,j}$ depends only on $x_{i.j}$, $y_{i,j}$ and $\hat{x}_{i,j}$. Given the knowledge of how to compute $f_{i,j}$, a straightforward, yet general, way to quantify the overall improvement in image fidelity is the weighted sum of $f_{i,j}$. In other words, the quantity m that is given by

$$m = \sum_{i,j} w_{i,j} f_{i,j}, \tag{2}$$

where $w_{i,j}$ is the weighting coefficient, provides the measurement of restoration performance. It can be shown that the SNR improvement (before expressed in dB) is a special case of the above formulation, with $f_{i,j} = (x_{i,j} - y_{i,j})^2/(x_{i,j} - \hat{x}_{i,j})^2$ and $w_{i,j} = (x_{i,j} - \hat{x}_{i,j})^2/\sum_{k,l}(x_{k,l} - \hat{x}_{k,l})^2$.

Based on formulation (2), we propose an improved measure by introducing a new metric for pixel fidelity improvement and incorporating main properties of the human visual system into the weighting coefficients. However, for comparative evaluation of performance measures, some basic requirements and evaluation criteria are needed to be addressed first.

2.2 Evaluation Criteria

Suppose the performance of a restoration is evaluated with two measuring processes, say M and M'. The question to be explored here is how good measure M is, as compared with measure M'. Obviously, accuracy and precision are of great importance in comparative evaluation of measures. Moreover, quality of the message conveyed from the measurement is worth being evaluated. In the following part of this section, these three important aspects of measure will be elaborated.

A. Accuracy of the measure

For a particular restoration (a particular set of x, y and $\hat{x}$), let m and m' be two measurements obtained with two different measures. The question that whether m or m' is more accurate can be answered only when we have an idea on what the 'true' value is, or when we have some criteria to evaluate their closeness to the true value. As for restoration performance measure, it is widely accepted that, the more closely a measure can mirror the subjective judgment made by human observers, the more accurate can be its measurements. However, an 'absolutely' accurate image quality measure that can be used as a representative of the visual system's response is still not yet available in the field. Therefore, the accuracy of an restoration performance measure is hard to be evaluated. It is, nevertheless, certain that an performance measure which incorporates properties of the human visual system should be more accurate than those which are solely based on MSE criterion.

B. Precision of the measure

Precision is an expression of relative smallness of variability in a measuring process [5]. A restoration performance measure of high precision, when applied to a series of *homogeneous restorations*, should produce a set of measurements of small spread. The smaller the spread of the measurements, the more precise will be the measure. Here, the meaning of 'homogeneous restoration' should be explained as clearly as possible to avoid ambiguity. Firstly, the meaning of 'a restoration' is in general an operation on a distorted version of a x, which is named as y, to obtain a $\hat{x}$. How good of that restoration means how much the image fidelity is improved from y to $\hat{x}$ with respect to x. Therefore, when we mention 'a restoration', we are referring to a particular set of x, y and $\hat{x}$. Two restorations are homogeneous if they have very similar x, y and $\hat{x}$. Of course, 'similar' is a concept requiring human observers to become meaningful. For unambiguity, we give a clearer definition here: Two restorations are said to be homogeneous if (i) their x's are of same texture, pattern and structure, (ii) their y's contain same type of distortion, and (iii) the restoration processes that produce $\hat{x}$'s are the same.

It will be shown in Section 4 that the SNR improvement may yield measurements of large spread when it is applied to a series of homogeneous restorations. This indicates that the SNR improvement is of low precision. This fact initiates our motivation to devise a more precise performance measure.

C. Meaning of the measurement

As a good restoration performance measure, its measurements should convey message of significant meaning to its users. Particularly, whether the image fidelity is improved or not after restoration should be revealed clearly. Moreover, the extent to which a restoration operator improves the image fidelity, as compared with the ideal restoration, should be conveyed through the measurement. Obviously, in a restoration, the amount of achievable image fidelity improvement has its maximum and minimum points. It makes no sense to have infinity improvement or deterioration in fidelity. A desirable measure should provide a finite positive value, say U, to indicate the maximally achievable improvement, and a finite negative value, say L, to indicate the maximally achievable deterioration. The measure should also report zero value when there is no improvement nor deterioration in image fidelity. As for other restoration results, the measurements should lie in the range $[L, U]$.

It can be seen that the SNR improvement contains only one reference point in it, namely zero, which is to indicate 'no improvement in SNR'. Thus a positive SNR improvement indicates that the image fidelity is improved, while a negative one indicates deterioration. However, the minimum of SNR improvement is at $-\infty$ and it happens when there is no distortion in y ($y = x$). It does not correspond to 'the maximally achievable deterioration'. Furthermore, the SNR improvement will encounter problems in interpreting the amount of fidelity improvement when one of the following three cases happens.

1. When $\hat{x} = y = x$, the SNR improvement is undefined.

2. When $\hat{x} = x$ and $y \neq x$, SNR Improvement $= \infty$. In other words, when the restored image is exactly the same as the original image, the SNR improvement will be of the same value (∞) no matter how large the distortion of the observed image is. However, in view of the amount of fidelity that is improved, this should not be the case since the fidelity improved with respect to an image with more distortion should be larger.

3. When $y = x$ and $\hat{x} \neq x$, SNR Improvement $= -\infty$. That is, no matter how close the restored image is to the original one, the SNR improvement will be of the same value.

Hence, the SNR improvement can not act as a good restoration performance measure, for that it convey unclear and, sometimes, even wrong message on the image fidelity improvement in a restoration.

3 An Improved Measure

As mentioned in previous section, we propose to apply a weighted sum of the *pixel fidelity improvement* as a measure of image restoration performance. Obviously, different definitions of pixel fidelity improvement and different choices of the weighting coefficients result in performance measures of different properties. As we have emphasized, our goal is to devise a measure which is more precise than the SNR improvement and whose measurements convey better quality message to its users. To incorporate these improvements, there are some constraints for the definition of pixel fidelity improvement and some factors that should be considered in choosing the weighting coefficients. In the following part of this section, we will address these two issues and then present our solutions.

3.1 A metric for pixel fidelity improvement

Firstly, the function that defines the pixel fidelity improvement $f_{i,j}$ should meet the following requirements.

1. It must be well-defined for any value of $x_{i,j}$, $y_{i,j}$ and $\hat{x}_{i,j}$;

2. The range of $f_{i,j}$ should be finite and include both positive and negative values (The improvement is negative when fidelity is deteriorated. The finite-range requirement is for that it makes no sense to have infinite improvement in fidelity.);

3. The maximum, zero and minimum of $f_{i,j}$ should respectively correspond to 'maximally achievable improvement', 'zero improvement' and 'maximally achievable deterioration'.

Consider a pixel at location (i, j), the error before restoration is $|x_{i,j} - y_{i,j}|$ and the error after restoration is $|x_{i,j} - \hat{x}_{i,j}|$. The fidelity of this pixel will be either improved (when $|x_{i,j} - y_{i,j}| > |x_{i,j} - \hat{x}_{i,j}|$) or deteriorated (when $|x_{i,j} - y_{i,j}| < |x_{i,j} - \hat{x}_{i,j}|$), and the amount of improvement (or deterioration) is $|x_{i,j} - y_{i,j}| - |x_{i,j} - \hat{x}_{i,j}|$. The maximally achievable amount of improvement is

$|x_{i,j} - y_{i,j}| - |x_{i,j} - x_{i,j}|$, and the maximally achievable amount of deterioration is $|x_{i,j} - y_{i,j}| - |x_{i,j} - z_{i,j}|$, where z is given by

$$z_{i,j} = \begin{cases} G, & x_{i,j} < G - x_{i,j} \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

Here, G is the maximum grey level ($G = 255$ for image of 256 grey level). Note that $z_{i,j}$ is never equal to $x_{i,j}$.

We propose to quantify the pixel fidelity improvement as the normalized amount of improvement, where normalization is done with respect to its maximally achievable amount of improvement. For clarity purpose, our definition of pixel fidelity improvement is denoted as $F_{i,j}$ and is given by

$$F_{i,j} = \begin{cases} \frac{|x_{i,j}-y_{i,j}|-|x_{i,j}-\hat{x}_{i,j}|}{|x_{i,j}-y_{i,j}|-|x_{i,j}-x_{i,j}|}, & |x_{i,j} - y_{i,j}| > |x_{i,j} - \hat{x}_{i,j}| \\ 0, & \text{if} \quad |x_{i,j} - y_{i,j}| = |x_{i,j} - \hat{x}_{i,j}| \\ -\frac{|x_{i,j}-y_{i,j}|-|x_{i,j}-\hat{x}_{i,j}|}{|x_{i,j}-y_{i,j}|-|x_{i,j}-z_{i,j}|}, & |x_{i,j} - y_{i,j}| < |x_{i,j} - \hat{x}_{i,j}| \end{cases} \tag{4}$$

One can see that $F_{i,j}$ is well-defined for any values of $\hat{x}_{i,j}$, $y_{i,j}$ and $x_{i,j}$. In particular, $F_{i,j} = 0$ if and only if $\hat{x}_{i,j} = y_{i,j}$; $F_{i,j} = 1$ if and only if $\hat{x}_{i,j} = x_{i,j}$ and $y_{i,j} \neq x_{i,j}$; and $F_{i,j} = -1$ if and only if $\hat{x}_{i,j} = z_{i,j}$ and $y_{i,j} \neq z_{i,j}$. When $\hat{x}_{i,j} = y_{i,j} = x_{i,j}$, $F_{i,j} = 0$. Moreover, $F_{i,j}$ is always within the range $[-1, 1]$.

3.2 Determination of weighting coefficients

Secondly, the weighting coefficients $w_{i,j}$ should be used to weight the relative influence of $F_{i,j}$ on image's fidelity, or in other words, the relative degree to which each pixel fidelity improvement affects the overall image fidelity. The influence of $F_{i,j}$ can be strong or weak, depending on the local characteristics of pixel (i, j) in the original image. In view of this, we first partition the image into several segments of different properties, and determine a weight for each segment based on its size and features. For each $F_{i,j}$, its weight, $w_{i,j}$, is then determined according to the segment to which the pixel (i, j) belongs.

In the segmentation, the image is first divided into two partitions according to the sign of $F_{i,j}$. This implies that one of them contains pixels with fidelity improvement and another contains pixels with deterioration. Each of these two partitions is further segmented into low-spatial-activity (or level) and high-spatial-activity (or edge) segments by thresholding the local variance of the pixel intensity in the original image. Consequently, the image is partitioned into four distinct segments, where each can be described by two attributes, namely, either deterioratedness (D) or improvedness (I), and either low-spatial-activity (L) or high-spatial-activity (H). For notational convenience, the four image segments are denoted as $R_{DL} = \{(i,j) : F_{i,j}{<}0 \text{ and } M_{i,j}{\le}t\}$, $R_{DH} = \{(i,j) : F_{i,j}{<}0 \text{ and } M_{i,j}{>}t\}$, $R_{IL} = \{(i,j) : F_{i,j}{\ge}0 \text{ and } M_{i,j}{\le}t\}$ and $R_{IH} = \{(i,j) : F_{i,j}{\ge}0 \text{ and } M_{i,j}{>}t\}$. Here, $M_{i,j}$ is the local variance of intensity at location (i,j), which is defined by [6]

$$M_{i,j} = \frac{1}{(2P+1)(2Q+1)} \sum_{m=i-P}^{i+P} \sum_{n=j-Q}^{j+Q} (x_{m,n} - \bar{x}_{i,j})^2, \tag{5}$$

where $(2P+1){\times}(2Q+1)$ is the extent of the analysis window and $\bar{x}_{i,j}$ is the local mean of $x_{i,j}$ over the analysis window. (In our realization, $P = Q = 1$.) The threshold value t is established as $t = 10^{\frac{1}{2}\log M_{max}}$, where M_{max} is the maximum local variance.

By making use of this segmentation scheme, the proposed measure, which is named as *Restoration Score* for the sake of reference, is given explicitly as

$$\text{Restoration Score} = \sum_{i,j} w_{i,j} F_{i,j} = \sum_{k\in\Omega} w_k N_k \bar{F}_k = \sum_{k\in\Omega} W_k \bar{F}_k, \tag{6}$$

where w_k is the common weight for every pixel in R_k, $\bar{F}_k$ is the mean of $F_{i,j}$ over R_k, N_k denotes the total number of pixels in R_k, $W_k = w_k N_k$ and $\Omega = \{DL, DH, IL, IH\}$.

In this work, two contributing factors are considered to determine W_k. One of them is the size and another is the features of the corresponding segment. These two factors are independent to each other and, hence, each weight W_k can be expressed as

$$W_k = W_{k,size} \cdot W_{k,feature}, \tag{7}$$

where $W_{k,size}$ and $W_{k,feature}$ are the shares based on the size and features of R_k respectively. In the following, we will discuss these two factors one by one separately.

A. Size Factor

It is postulated that human observer will not monitor and process an image pixel by pixel. Rather than that, the human observer tends to divide the whole image into several regions, such as edge and level regions, and then evaluates the image region by region [7]. In the light of this, low-spatial-activity segment and high-spatial-activity segment are considered separately in evaluating the $W_{k,size}$.

For the low-spatial-activity segment, whether $\bar{F}_{IL}$ or $\bar{F}_{DL}$ is more significant depends on the proportion of improvement to deterioration: In the beginning, when the improvement (deterioration) proportion is very small, it is insignificant to the human viewer. The significance increases rapidly when the proportion increases. Finally, when the proportion becomes large enough, the increase of the significance slows down. In order to incorporate this property, $W_{IL,size}$ and $W_{DL,size}$ are determined respectively as $S(L_I)$ and $S(L_D)$, where $L_I = \frac{N_{IL}}{N_{IL}+N_{DL}}$, $L_D = \frac{N_{DL}}{N_{IL}+N_{DL}}$ and S is a S-curve-like monotonic increasing function. Note that L_I and L_D are counterproductive in a way that $L_I + L_D = 1$. Moreover, for the sake of normalization, it is desirable for S to have the properties that (i) $S(0) = 0$,$S(1) = 1$ and (ii) $S(t) + S(1-t) = 1$ for $t \in [0, 1]$. The following function is devised for the required purpose.

$$S(t) = \begin{cases} \frac{(2t)^3}{2}, & 0 \le t \le 0.5 \\ 1 - \frac{[2(1-t)]^3}{2}, & 0.5 < t \le 1 \end{cases} \quad (8)$$

The plot of S is shown in Figure 1.

The same philosophy applies to the high-spatial-activity segment. Specifically, $W_{IH,size} = S(H_I)$ and $W_{DH,size} = S(H_D)$, where $H_I = \frac{N_{IH}}{N_{IH}+N_{DH}}$ and $H_D = \frac{N_{DH}}{N_{IH}+N_{DH}}$.

B. Feature Factor

Another important property of the human visual system is that the noise in image regions of low spatial activity is more visible than that in regions of high spatial activity, which is referred to as *spatial visual masking* [7-8]. It is therefore desirable to give more weight to $\bar{F}_{DL}$ than $\bar{F}_{DH}$ in a way that $W_{DL,feature} > W_{DH,feature}$.

Since high-frequency components of an image are typically destroyed in the blurring process, the distortions in high-spatial-activity regions are larger than those in low-spatial-activity regions.

In that case, the purpose of image restoration is mainly to reconstruct the high-frequency components of the image [9]. Therefore, the improvement in high-spatial-activity segment is more important than that in low-spatial-activity segment. This implies that more weight should be given to $\bar{F}_{IH}$ than $\bar{F}_{IL}$, i.e. $W_{IH,feature} > W_{IL,feature}$.

Unlike $W_{k,size}$'s, which vary according to the size of their corresponding segment, $W_{k,feature}$'s are constants that must be pre-determined in the measurement. For the purpose of normalization, the values of $W_{k,feature}$'s are chosen in a way that $W_{DL,feature} + W_{DH,feature} = 1$ and $W_{IL,feature} + W_{IH,feature} = 1$.

We have simulated the proposed measure with different values of $W_{k,feature}$'s to select their appropriate combination. In particular, an experiment in which human viewers were asked to rank the quality of a series of restored images was conducted. These restored images were generated from a testing image that undergoes various distortions and then a restoration. This experiment was repeated many times with different testing images and different restoration operators. To minimize the error between subjective ranking and the proposed measure [10], we found that $W_{DL,feature} = 0.8$ and $W_{IH,feature} = 0.9$ (hence $W_{DH,feature} = 0.2$ and $W_{IL,feature} = 0.1$) could provide an adequately good representation as perceived by human viewers.

3.3 Properties of the proposed measure

Let us consider the following three special cases.

1. When $\hat{x} = x$ and $|x_{i,j} - y_{i,j}| > |x_{i,j} - \hat{x}_{i,j}|$ for all $\hat{x}_{i,j}$, we have Restoration Score = 1.

2. When $\hat{x} = y$, we have Restoration Score = 0.

3. When $\hat{x} = z$ and $|x_{i,j} - y_{i,j}| < |x_{i,j} - \hat{x}_{i,j}|$ for all $\hat{x}_{i,j}$, we have Restoration Score = -1.

One can see that the measurements $+1$, 0 and -1 respectively correspond to 'the maximally achievable improvement', 'no improvement' and 'the maximally achievable deterioration'. Any value of Restoration Score is confined to the range $[-1, +1]$. Unlike the SNR improvement, Restoration Score contains three finite reference points, which are useful in providing its users with a better insight into the effectiveness of the restoration method being evaluated.

In the performance measurement by means of the SNR improvement, each pixel is considered to be equally important. Thus the SNR improvement of a restoration is very sensitive to the content of the original image, rather than the distortion introduced and the restoration applied. However, the Restoration Score takes account of local properties of the image and treats each pixel differently. Since image segmentation is performed in the measurement, the restoration score is less sensitive to the content of the image and depends mainly on the distortion and the restoration operator. An example is provided to shed light on this property: Figure 2 shows a series of three homogeneous restorations which are obtained with three similar testing images, same type of distortion and same type of restoration. Their SNR improvements are respectively -3.84, -1.65 and -2.62, while their Restoration Scores are all -0.05. The image-content-insensitive property of the Restoration Score make it more precise in measuring restoration performance. Relative precision of the Restoration Score will be evaluated and compared with the SNR improvement in next section.

Finally, we remark that the fore-mentioned properties of the Restoration Score are invariant to the choices of $W_{k,feature}$'s. The values of $W_{k,feature}$'s will only affect the accuracy of the Restoration Score with respect to the human visual system. Since $W_{k,feature}$'s are constants in computing the Restoration Score, they will not affect the variability in the measurement, i.e the precision.

4 Precision Evaluation of the Measure

Precision is generally measured with the standard deviation of the measurements obtained from the same homogeneous materials. However, when two measures are of different units, their precision cannot be compared directly by using their corresponding standard deviations. Hence, we adopt a criterion described in literature [5] to evaluate the precision. Let u and v represent the measurements of two competing methods. Assume that the curve of v versus u is linear over a small range. Consider a particular value of u, say u_0, and the corresponding value of v is v_0. Let the standard deviation of u and v near the point (u_0, v_0) be σ_u and σ_v respectively, and the slope at $u = u_0$ be represented by $\Delta v/\Delta u$. Then, a criterion used to reveal the precision of v over u is given by the quantity

$$S_{vu} = \frac{\Delta v/\Delta u}{\sigma_v/\sigma_u}, \qquad (9)$$

which is call the *sensitivity of* v *with respect to* u. That v is more precise than u is indicated by a value of S_{vu} larger than unity, and vice verse. It is important to note that the sensitivity S is a local quantity: both σ_u and σ_v may be functions of the level of measurement, and the slope $\Delta v/\Delta u$ may vary from one point to another. Therefore, the sensitivity S should be expressed as a function of measurement levels.

A number of experiments were carried out to compare the precision of the proposed measure with that of the SNR improvement. The sensitivity criterion was applied as a comparison vehicle for the precision evaluation of the two measures.

A. Experiment 1

In the first experiment, a set of eight similar images, which are depicted in Figure 3, were used as the original images. Eight homogeneous restorations were generated by introducing same distortion to the original images, and then applying same restoration operator to obtain the restored images. In order to obtain different sets of homogeneous restorations, we varied the distortion added. We exploited totally 120 different distortions. Each was a combination of one of the four blurs described in Appendix, and White Gaussian noise at one of the 30 various levels, ranged from 1 to 30 dB BSNR. All the distorted images were then restored with the Wiener filter [9].

For each set of homogeneous restorations, we computed its i) average SNR improvement, ii) standard deviation of SNR improvement, iii) average Restoration Score, and iv) standard deviation of Restoration Score. The 120 averages of the SNR improvements were plotted against their corresponding averages of Restoration Scores in Figure 4. We then applied the least-squares method to fit a polynomial curve through these data points. The computed curve is also shown in Figure 4. For the i-th data point, the tangent at this point was determined, and the sensitivity of Restoration Score with respect to the SNR improvement, denoted as S_R, was then computed with

$$(S_R)_i = (\text{tangent of the curve})_i \cdot \frac{(\text{s.d. of SNR Improvement})_i}{(\text{s.d. of Restoration Score})_i}. \tag{10}$$

The sensitivity S_R at each data point is plotted against its corresponding Restoration Score in Figure 5. It is shown that S_R is well above unity when Restoration Score ranges from -0.4 to 0.7. This shows that Restoration Score, in this measurement range, is much more precise than the SNR improvement in measuring restoration performance.

B. Experiment 2

In this experiment, eight natural text images, which are depicted in Figure 6, were used as the original images. 120 sets of homogeneous restorations were obtained by varying the distortions added to the originals, which were similar to the first experiment, and then restoring the distorted images with the iterative restoration algorithm [3]. The averages of the SNR improvements were plotted against their corresponding averages of Restoration Scores in Figure 7. By following the procedures described in Experiment 1, we compute the S_R for each set of homogeneous restorations. The results were plotted against their corresponding Restoration Scores in Figure 8. It is found that S_R is well above unity at measurement levels ranged from 0.0 to 0.6 of Restoration Score, and it thus justifies once again that the precision of Restoration Score is higher than that of the SNR improvement.

C. Experiment 3

In the simulations presented above, we have shown that the precision of Restoration Score, in a range from -0.4 to 0.7, is higher than that of the SNR improvement. In terms of SNR improvement, this is corresponding to a range of 20dB from -10dB to 10dB. In practical situation of interests, this range of measurements can cover all possible results obtained with any realistic restoration. To make the precision evaluation more complete, however, we also explore the precision of Restoration Score at measurement levels outside the range [-0.4, 0.7].

In the first part of this experiment, we exploited 4 hypothetical restoration operators to produce restored images that were close to the original image. The outputs of these restoration operators are simply $x+\eta$, where η is a zero-mean white noise signal with a variance σ_η^2 of either 0.5, 1.0, 1.5 or 2.0. The testing images depicted in Figure 3 were applied. We first obtained 240 sets of homogeneous restorations with 60 different distortions and the above 4 restoration operators. Applying similar procedures in the previous experiments, we obtained the results given in Figure 9. It shows that the sensitivity is well above unity from 0.7 to 1.0. Hence, the Restoration Score was found to be more precise than the SNR improvement at that range of measurements.

In the second part of this experiment, same set of testing images were applied, but this time we exploited another set of hypothetical restoration operators to produce images that were close to the one with maximum deterioration. The outputs of these operators are $z+\mu$, where z is the image defined by (2), and μ is a zero-mean white noise with variance σ_μ^2. By varying σ_μ^2 ($\sigma_\mu^2 = 0.5, 1.0, ...,$ and 5.0), 10 restoration operators were formed. Together with 24 different distortions, we obtained 240 sets of homogeneous restorations. The results we obtained, which are plotted in Figure 10, show that the precision of Restoration Score is higher than that of the SNR improvement for the measurement level around -1.0.

5 Conclusions

In this paper, we have proposed a novel quantitative measure of image restoration performance. It has been shown by detailed experiments that the proposed measure is more precise than the SNR improvement. Another feature of the proposed measure is that it contains clearly-defined and meaningful reference points in its measurements. These reference points are useful in providing users with a better insight into the effectiveness of the restoration algorithm under study. The proposed measure overcomes main limitations of the SNR improvement, and it can be a better performance index for the comparative evaluation of image restoration algorithms. Note that the proposed measure is by no means optimal as finding an optimal measure is currently limited by our immature knowledge of human visual behavior and the knowledge of how to incorporate the known properties of the human visual system into the measurement.

Appendix

Description of Distortions Added to Originals

The following blurs were exploited in the experiments.

1) Out-of-focus blur with the following PSF model [11]:

$$d(i,j) = \frac{1}{20.0296}\begin{bmatrix} 0.1716 & 0.7929 & 1.0000 & 0.7929 & 0.1716 \\ 0.7929 & 1.0000 & 1.0000 & 1.0000 & 0.7929 \\ 1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 \\ 0.7929 & 1.0000 & 1.0000 & 1.0000 & 0.7929 \\ 0.1716 & 0.7929 & 1.0000 & 0.7929 & 0.1716 \end{bmatrix} \qquad (A1).$$

2) Out-of-focus blur with the following PSF model [12]:

$$d(i,j) = \begin{cases} 1/57, & i^2 + j^2 \leq 17 \\ 0, & \text{otherwise.} \end{cases} \qquad (A2).$$

3) Linear motion blur with the following PSF model [3]:

$$d(i) = \begin{cases} 1/9, & i = 0, 1, \ldots, 8 \\ 0, & \text{otherwise.} \end{cases} \qquad (A3).$$

4) Linear motion blur with the following PSF model [3]:

$$d(i) = \begin{cases} 1/15, & i = 0, 1, \ldots, 14 \\ 0, & \text{otherwise.} \end{cases} \qquad (A4).$$

The original images were first distorted by one of the above blurs, and white Gaussian noise was then added at various BSNR (blurred signal-to-noise ratio) defined as

$$\mathrm{BSNR} = 10 \log \left(\frac{\text{blurred image variance}}{\text{noise variance}} \right) \qquad (A5).$$

References

[1] M. I. Sezan and A. M. Tekalp, 'Survey of recent developments in digital image restoration,' *Opt. Eng.* 29(5), 393-404 (1990).

[2] A. K. Katsaggelos (ed.), *Digital Image Restoration,* pp. 17-18, Springer-Verlag, Berlin, New York (1991).

[3] R. L. Lagendijk and J. Biemond, *Iterative Identification and Restoration of Images,* Kluwer Academic Publishers, Boston (1991).

[4] A. K. Katsaggelos, 'Iterative image restoration algorithms,' *Opt. Eng.* 28(7) 735-748 (1989).

[5] John Mandel, *Evaluation and Control of Measurements,* Marcel Dekker, Inc., New York (1991).

[6] J. S. Lee, 'Digital image enhancement and noise filtering by use of local statistics,' *IEEE Trans. Pattern Analysis Mach. Intel.* PAMI-2(2), 165-168 (1980).

[7] J. O. Limb, 'Distortion criteria of the human viewer,' *IEEE Trans. Syst., Man, Cyber.* SMC-9(12), 778-793 (1979).

[8] G. L. Anderson and A. N. Netrarali, 'Image restoration based on a subjective criterion,' *IEEE Trans. Syst., Man, Cyber.* SMC-6(12), 845-853 (1976).

[9] H. C. Andrews and B. R. Hunt, *Digital Image Restoration,* Prentice Hall, Englewood Cliffs, N.J. (1977).

[10] W. K. Pratt, 'Sampled image quality measures,' Chap. 7 in *Digital Image Processing,* pp. 162-198, John Wiley & Sons, New York (1978).

[11] R. L. Lagendijk, J. Biemond and D. E. Boekee, 'Regularized iterative image restoration with ringing reduction,' *IEEE Trans. Acoustics, Speech, Signal Processing* ASSP-36(12) 1874-1887 (1988).

[12] S. J. Reeves, 'Optimal space-varying regularization in iterative image restoration,' *IEEE Trans. Image Processing* IP-3(3), 319-324 (1994).

List of Figure Captions

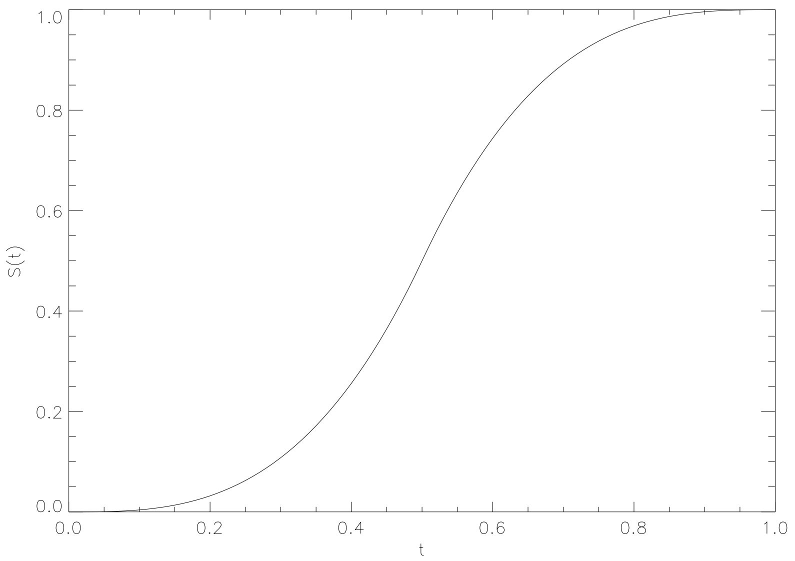

Figure 1: Plot of the S-function

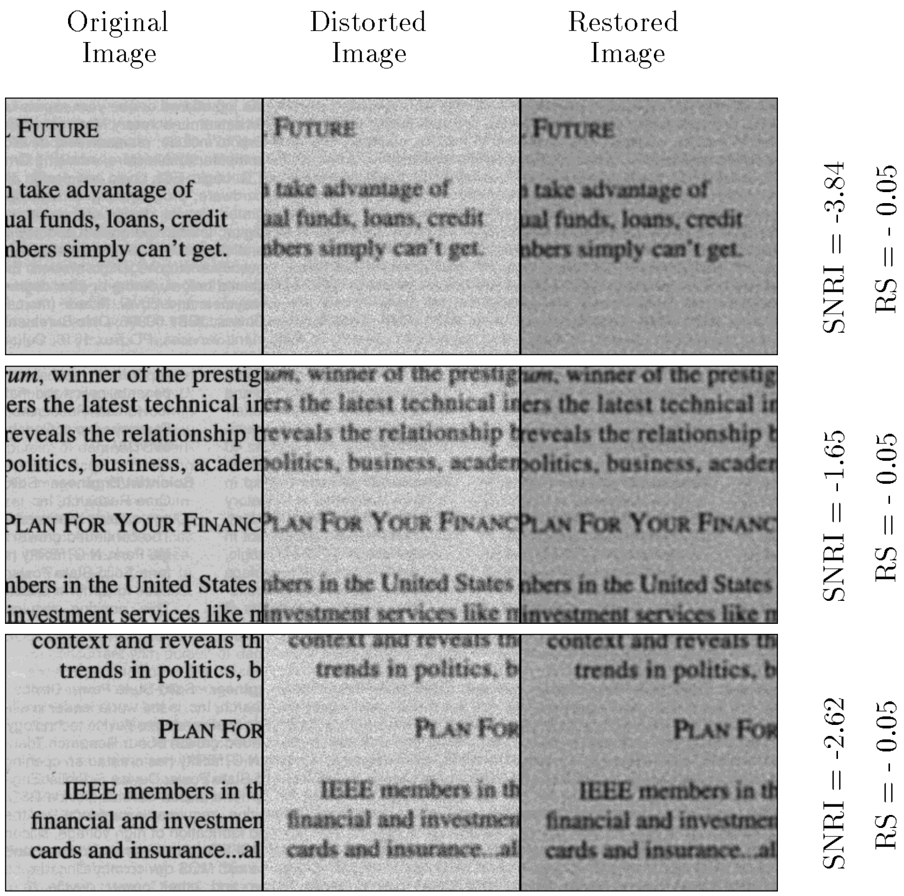

Figure 2: Example that illustrates the image-content-insensitive property of the Restoration Score.

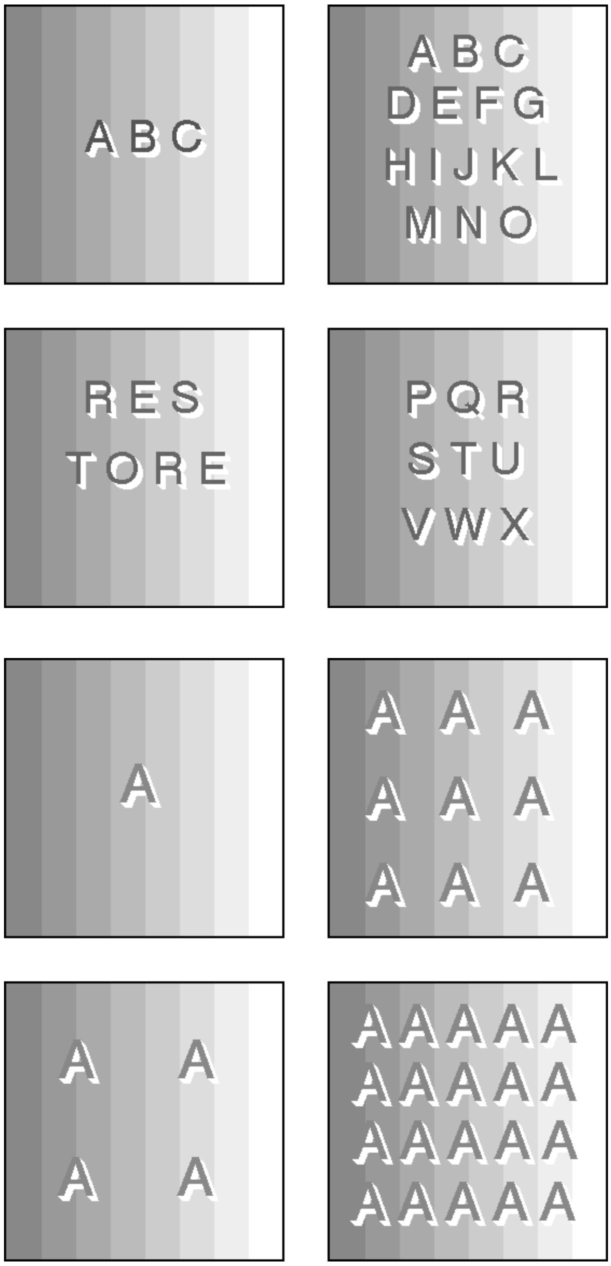

Figure 3: Original testing images to which distortions were introduced in Experiment 1.

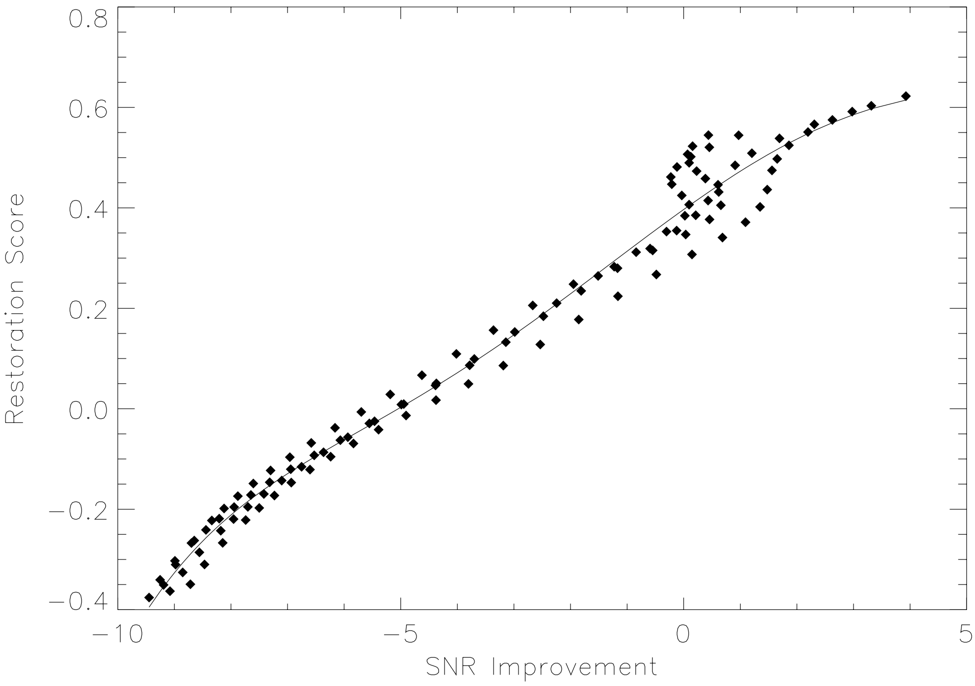

Figure 4: Plot of the average Restoration Score versus the average SNR improvement for 120 sets of homogeneous restorations in Experiment 1.

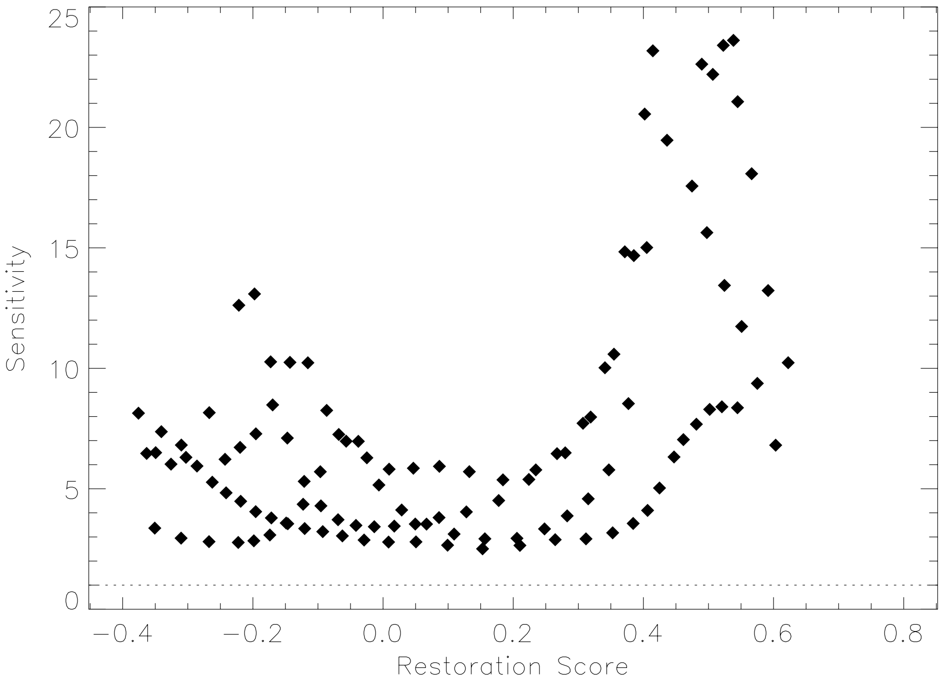

Figure 5: Plot of S_R (the sensitivity of Restoration Score with respect to SNR improvement) against Restoration Score for 120 sets of homogeneous restorations in Experiment 1.

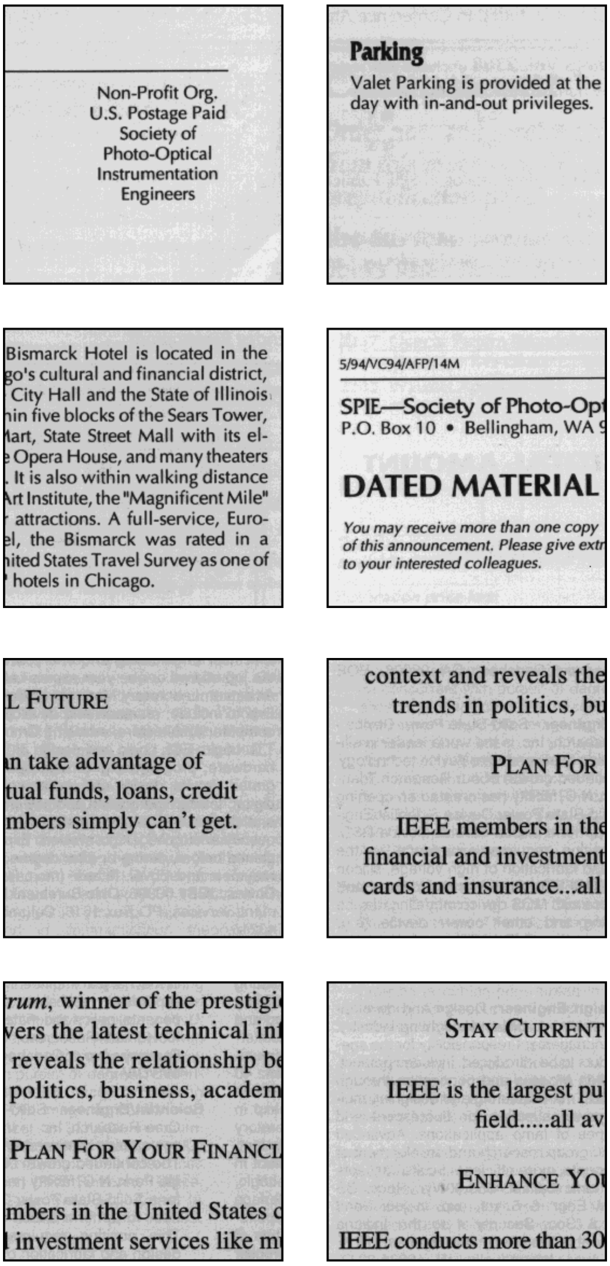

Figure 6: Original text images to which distortions were introduced in Experiment 2.

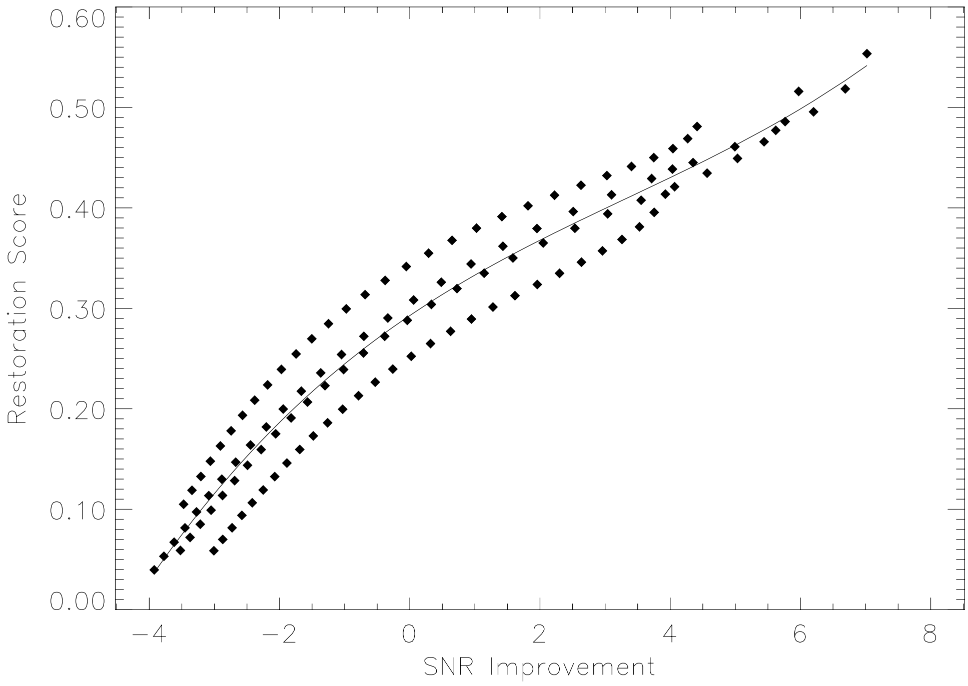

Figure 7: Plot of the average Restoration Score versus the average SNR improvement for 120 sets of homogeneous restorations in Experiment 2.

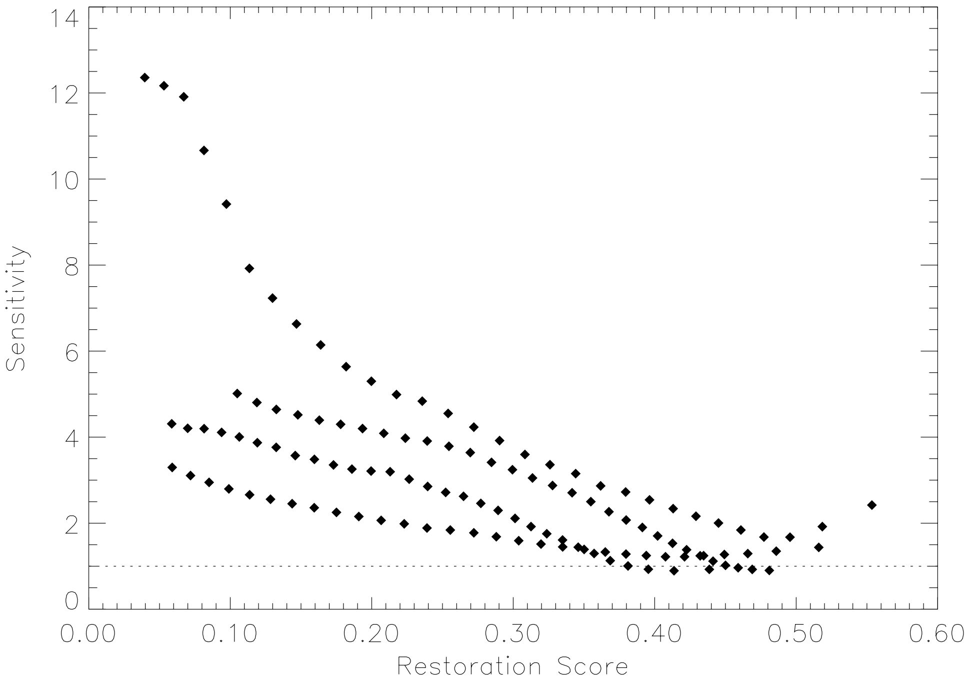

Figure 8: Plot of S_R (the sensitivity of Restoration Score with respect to SNR improvement) against Restoration Score for 120 sets of homogeneous restorations in Experiment 2.

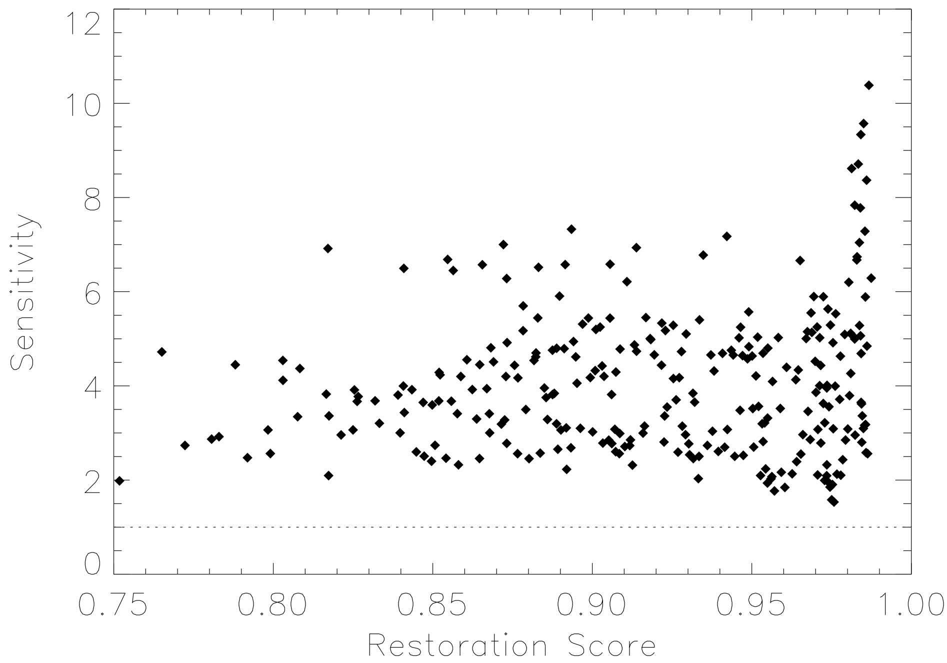

Figure 9: Plot of S_R (the sensitivity of Restoration Score with respect to SNR improvement) against Restoration Score for 240 sets of homogeneous restorations in the first part of Experiment 3.

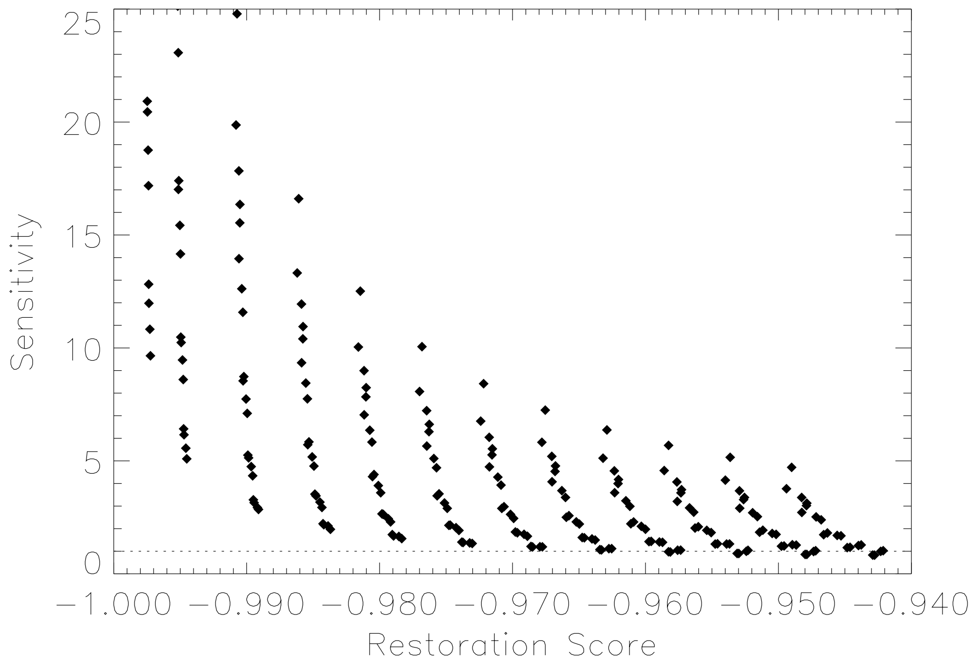

Figure 10: Plot of S_R (the sensitivity of Restoration Score with respect to SNR improvement) against Restoration Score for 240 sets of homogeneous restorations in the second part of Experiment 3.