AN IMPROVED QUANTITATIVE MEASURE OF IMAGE RESTORATION QUALITY

Steven S. O. Choy, Yuk-Hee Chan and Wan-Chi Siu

Department of Electronic Engineering
The Hong Kong Polytechnic University
Hung Hom, Kowloon, Hong Kong

ABSTRACT

The field of image restoration lacks promising comparison vehicle for judging the effectiveness of competing algorithms. By far the most widely adopted quantitative measure of image restoration quality is the SNR improvement. However, we find that the SNR improvement is of low precision, which will adversely hinder it from being a reliable measure. It is also noted that another limitation of the SNR improvement is that it cannot reveal clearly the extent to which the image quality is improved. In this paper, we devise an alternative measure for quantitative evaluation of image restoration quality. The proposed measure is much more precise than the SNR improvement. Moreover, the proposed measure contains finite and meaningful reference points in its measurements, to provide us with a better insight into the effectiveness of restoration algorithms that under study.

1. INTRODUCTION

Image restoration refers to the problem of reconstructing or estimating the uncorrupted 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. Many new restoration algorithms, ranging from deterministic iterative methods to optimal stochastic filtering, have been proposed in the last decade [1]. While many researchers have devoted considerable effort to the problem of image restoration, few studies have been undertaken for performance evaluation of restoration methods. One of the underlying reasons is that accurate image quality measure for restoration is hardly available. The field of image restoration lacks promising criterion to evaluate the performance of the algorithms [2].

In applications of image restoration, image quality usually refers to the image's fidelity to its original. In this paper, we use the term 'image restoration quality' to refer to the effectiveness of restoration in improving the fidelity of the processed image. To measure the image restoration quality thus means to measure the amount of improvement in image quality due to restoration. By far the most popular quantitative measure of image restoration quality is the SNR improvement, which is defined as the difference between the SNR of the restored image and the SNR of the noisy blurred image, or mathematically,

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

where x, y and $\hat{x}$ are the original image, the degraded image and the restored image, respectively. This objective measure is usually applied to evaluate restoration performance, and is widely adopted in the comparative study of restoration algorithms [2, 3].

However, we find that the SNR improvement has some limitations in measuring image restoration quality. In next section, the properties that a good measure of image restoration quality should have, as well as the limitations of the SNR improvement, will be discussed in detail. An improved measure of image restoration quality is then proposed. The derivation of the proposed measure 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 QUALITY

Suppose the image restoration quality is measured with two measuring processes, say M and M'. The question to be explored here is how good measure M is, when compared with measure M'. Obviously, accuracy and precision are of great importance in comparative evaluation of the two measures. Moreover, the 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

An accurate measure of image restoration quality should closely mirror the subjective judgment made by human observers. However, there exists no clear definition of image quality and an 'absolutely' accurate measure of image quality is still not yet available in the field. Therefore, there is no reliable criterion to evaluate the accuracy of an image restoration quality measure.

B. Precision of the measure

Precision is an expression of relative smallness of variability, or in other words, of relative stability within the measuring process [4]. Suppose a set of homogeneous distorted

Acknowledgment: This work was supported by the PolyU Research Grant: Project account 340.813.A3.420.

images were restored by the same restoration operator. A high precision measure of image restoration quality, when applied to these restored images, should produce a set of measurements of small spread. The smaller the spread of the measurements, the more precise will be the measure. Here, a set of homogeneous distorted images means a set of similar images of same type of blur and with same amount of noise. A set of homogeneous restored images is in turn a set of homogeneous distorted images restored by the same restoration operator.

C. Meaning of the measurement

As stated in (1), the SNR improvement is defined as the difference between SNRs of the images before and after restoration. A positive SNR improvement indicates that the quality of distorted image is improved, while a negative one indicates deterioration. The zero value of the SNR improvement indicates there is no improvement nor deterioration. However, when two different restoration methods are compared with each other by means of the SNR improvement, it only reveals which method is better. That how much it is better cannot be revealed clearly since the ideal SNR improvement is ∞. In view of this, it is desirable to define meaningful and finite reference points in a measure of image restoration quality. Obviously, the amount of image quality improvement in a restoration has its maximum and minimum points. A desirable measure should provide a finite positive value, say U, to indicate this maximally achievable improvement. On the other hand, the measure should provide 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 quality. As for other restoration results, the measurement should range from L to U. It is another goal of this paper to devise a measure having this property besides that of high precision.

In addition, there are problems in the SNR improvement when one of the following cases happens. i) The SNR improvement is undefined when $\hat{x} = y = x$; ii) When $\hat{x} = x$ and $y \neq x$, the SNR improvement will be of same value (∞) no matter how close the distorted image is to the original image; iii) When $y = x$ and $\hat{x} \neq x$, the SNR improvement will be of same value $(-\infty)$ no matter how close the restored image is to the original image. Hence, the SNR improvement cannot reveal correctly the image restoration quality in these three cases.

3. AN IMPROVED MEASURE

The purpose of image restoration is to improve the fidelity of a degraded image. To measure the fidelity improvement, the departure of the degraded image from the original image, as well as the departure of the restored image from the original image, is needed to be computed. Consider a pixel at location (i,j), the error of the image before restoration is $|x_{i,j} - y_{i,j}|$ and the error after restoration is $|x_{i,j} - \hat{x}_{i,j}|$. There will be either an improvement (when $|x_{i,j} - y_{i,j}| > |x_{i,j} - \hat{x}_{i,j}|$) or a deterioration (when $|x_{i,j} - y_{i,j}| < |x_{i,j} - \hat{x}_{i,j}|$), and the amount of improvement (or deterioration) is $|x_{i,j} - y_{i,j}| - |x_{i,j} - \hat{x}_{i,j}|$. For a certain pixel at location (i,j), its maximally achievable amount of improvement is $|x_{i,j} - y_{i,j}| - |x_{i,j} - x_{i,j}|$, and its maximally achievable amount of deterioration is $|x_{i,j} - y_{i,j}| - |x_{i,j} - z_{i,j}|$, where z is given by

$$z_{i,j} = \begin{cases} G, & x_{i,j} < G - x_{i,j} \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

Here, G is the maximum pixel value ($G = 255$ for image of 256 grey level). In general, the fidelity improvement of a certain pixel can be quantified as the normalized amount of improvement at that pixel, where normalization is done with respect to its maximally achievable amount of improvement. The fidelity improvement of each individual pixel is denoted as $F_{i,j}$ and is given by

$$F_{i,j} = \begin{cases} \frac{|x_{i,j}-y_{i,j}|-|x_{i,j}-\hat{x}_{i,j}|}{|x_{i,j}-y_{i,j}|}, & |x_{i,j} - y_{i,j}| > |x_{i,j} - \hat{x}_{i,j}| \\ 0, & |x_{i,j} - y_{i,j}| = |x_{i,j} - \hat{x}_{i,j}| \\ -\frac{|x_{i,j}-y_{i,j}|-|x_{i,j}-\hat{x}_{i,j}|}{|x_{i,j}-y_{i,j}|-|x_{i,j}-z_{i,j}|}, & |x_{i,j} - y_{i,j}| < |x_{i,j} - \hat{x}_{i,j}| \end{cases} \tag{3}$$

Having devised the concept on the fidelity improvement of each individual pixel, we propose to take the weighted sum of $F_{i,j}$'s as a measure of image restoration quality. To devise the weighting coefficient for $F_{i,j}$, each pixel is first classified into one of L classes according to its nature. A common weight w_k is then determined and assigned to every pixels in class R_k, $k = 1, 2, \ldots, L$. The proposed measure can therefore be formulated as

$$\sum_{k=1}^{L} \sum_{F_{i,j} \in R_k} w_k F_{i,j} = \sum_{k=1}^{L} W_k \bar{F}_k \tag{4}$$

where $\bar{F}_k$ is the mean of $F_{i,j}$ over R_k and $W_k = N_k w_k$ (N_k is the number of pixels in R_k).

For the sake of simplicity, the image pixels are classified into four classes only here. The image is first divided into two partitions according to the sign of $F_{i,j}$. That implies one of them contains pixels with improvement in fidelity and another contains pixels with deterioration. Each of these two partitions is further divided into two sub-partitions according to the local spatial activity of the pixels. For notational convenience, the four classes are expressed as $R_{IL} = \{(i,j) : F_{i,j} \geq 0 \textit{ and } M_{i,j} \leq t\}$, $R_{IH} = \{(i,j) : F_{i,j} \geq 0 \textit{ and } M_{i,j} > t\}$, $R_{DL} = \{(i,j) : F_{i,j} < 0 \textit{ and } M_{i,j} \leq t\}$, and $R_{DH} = \{(i,j) : F_{i,j} < 0 \textit{ and } M_{i,j} > t\}$. The parameter $M_{i,j}$ is the local variance of intensity at location (i,j), which is applied as a measure of spatial activity. The threshold t is given by $t = 10^{\frac{1}{2}\log M}$, where M is the maximum local variance of the image.

The proposed measure, which is named as *Restoration Score* for the sake of reference, can then be written explicitly as follows.

$$\begin{aligned} \text{Restoration Score} &= W_{IL}\bar{F}_{IL} + W_{IH}\bar{F}_{IH} \\ &\quad + W_{DL}\bar{F}_{DL} + W_{DH}\bar{F}_{DH} \end{aligned} \tag{5}$$

where W_{IL}, W_{IH}, W_{DL} and W_{DH} are the weights assigned to the corresponding class of pixels.

To determine the four weights, we take the following three factors into account.

i) The first factor is the proportion of improvement pixels to deterioration pixels. When the proportion of improvement pixels is very small, the overall fidelity improvement is insignificant to the human visual system. The significance increases rapidly when the proportion increases. Finally, when the proportion becomes large enough, the increase of the significance slows down.

Let N_{IL}, N_{IH}, N_{DL} and N_{DH} be the number of pixels in their corresponding classes. In the image partition of low spatial activity, the proportion of improvement pixels is $I_L = N_{IL}/(N_{IL}+N_{DL})$ and the proportion of deterioration pixels is $D_L = 1 - I_L$. Similarly, in the image partition of high spatial activity, the proportion of improvement pixels is $I_H = N_{IH}/(N_{IH}+N_{DH})$ and the proportion of deterioration pixels is $D_H = 1 - I_H$. In this proposed measure, the weights are determined according to the following: $W_{IL} \propto S(I_L)$, $W_{DL} \propto S(D_L)$, $W_{IH} \propto S(I_H)$, and $W_{DH} \propto S(D_H)$, where S is a function used to reflect the fore-mentioned properties of the human visual system and is devised as

$$S(t) = \begin{cases} \frac{(2t)^3}{2}, & 0 \le t \le 0.5 \\ 1 - \frac{[2(1-t)]^3}{2}, & 0.5 < t \le 1 \end{cases} \tag{6}$$

ii) Another important properties 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 [5], which is referred to as spatial visual masking. It is therefore desirable to give more weight to $\bar{F}_{DL}$ than $\bar{F}_{DH}$.

iii) Since high-frequency components of an image are typically destroyed in the blurring process, the distortion in the image regions of high spatial activity is larger than that in regions of low spatial activity. In typical application, the purpose of image restoration is mainly to reconstruct the high-frequency components of the image. Therefore, the improvement in image partition of high spatial activity is more important than that of low spatial activity. This implies that more weight should be given to $\bar{F}_{IH}$ than $\bar{F}_{IL}$.

Based on these three factors, the weights are chosen as

$$W_{IL} = (1-a) \cdot S(I_L) \tag{7}$$
$$W_{IH} = a \cdot S(I_H) \tag{8}$$
$$W_{DL} = b \cdot S(D_L) \tag{9}$$
$$W_{DH} = (1-b) \cdot S(D_H) \tag{10}$$

where a and b are positive constants. We have found experimentally that $a = 0.9$ and $b = 0.8$ provide an adequately good representation of measure in a way that it is monotonically related to the subjective measure.

Let's 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 $+1$ and -1 respectively correspond to the maximally achievable improvement and the maximally achievable deterioration. Any value of Restoration Score are confined to the range $[-1, +1]$. Unlike the SNR improvement, Restoration Score contains finite reference points, namely, 1, 0 and -1. These reference points are useful in providing its users with a better insight into the effectiveness of the restoration method being evaluated.

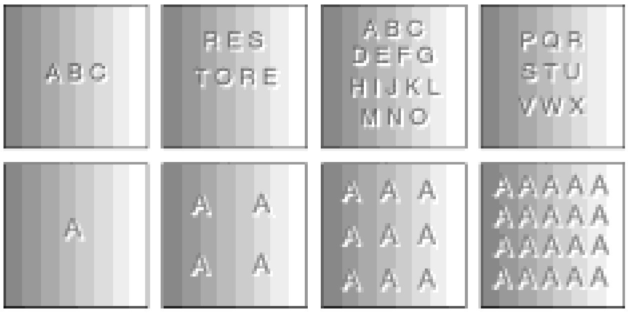

Figure 1: Original eight simulated images to which distortions were added.

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 precisions cannot be compared directly by using their corresponding standard deviations. Hence, we adopt a criterion described in literature [4] to evaluate the precisions. 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

$$S_{vu} = \frac{\Delta v/\Delta u}{\sigma_v/\sigma_u} \tag{11}$$

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 the level of measurement.

A number of experiments were carried out to compare the precision of the proposed measure with that of the SNR improvement. We present one of them in this section. In this experiment, the eight testing images depicted in Figure 1 were used. All these images have the same background but different letters of the same font in the foreground. In the first place, 120 sets of homogeneous distorted images were first generated by introducing 120 different distortions to all original images. The 120 different distortions were generated with four different blurs and 30 levels of white Gaussian noise ranged from 1 dB to 30 dB BSNR. All the distorted

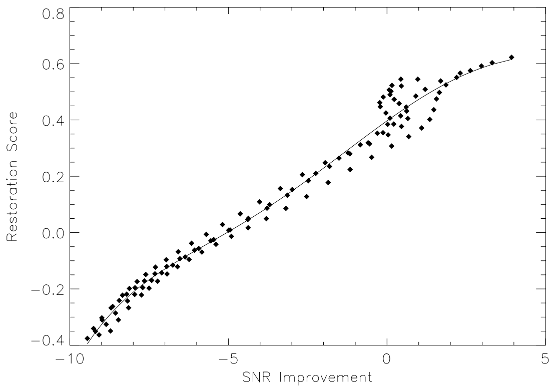

Figure 2: Plot of the average Restoration Score versus the average SNR improvement for 120 sets of homogeneous restored images.

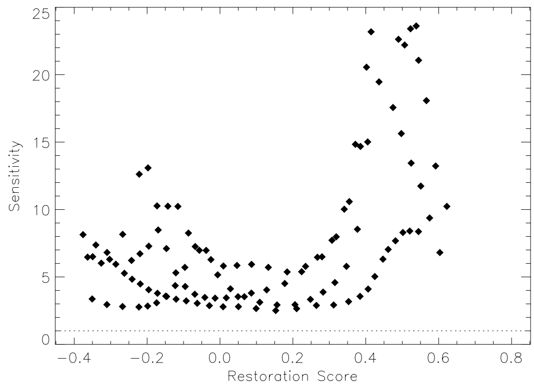

Figure 3: Plot of S_R (the sensitivity of Restoration Score with respect to SNR improvement) against Restoration Score for 120 sets of homogeneous restored images.

images were then restored with the use of the Wiener filter [6] to obtain 120 sets of homogeneous restored images. For each set of homogeneous restored images, 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 improvement were plotted against their corresponding averages of Restoration Score in Figure 2. We then applied the least-squares method to fit a polynomial curve through these data points. The computed curve is also shown in Figure 2. For the i-th data point, the slope 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

$$(S_R)_i = (\text{slope})_i \cdot \frac{(\text{s.d. of SNR Improvement})_i}{(\text{s.d. of Restoration Score})_i} \qquad (12)$$

The sensitivity at each data point is plotted against its corresponding Restoration Score in Figure 3.

It is found that S_R is well above unity when Restoration Score ranges from -0.40 to 0.65. This shows that Restoration Score, in this measurement range, is much more precise than the SNR improvement in measuring image restoration quality. Besides this experiment, we also conducted a number of similar experiments to evaluate the precision of Restoration Score. In these experiments, different sets of testing images were used and different restoration methods were applied to produce sets of homogeneous restored images for our analysis. Following the same procedure described before, we found that the precision of Restoration Score is higher than that of the SNR improvement.

In our simulations, the Restoration Score of the restoration results obtained ranged form -0.4 to 0.7. In terms of the SNR improvement, this is corresponding to a range of 20dB from -10dB to 10dB. In practice, this range of restoration results can cover all possible results obtained with any realistic restoration operator. To make the precision evaluation more complete, however, we also explored the precision at measurement levels outside the range [-0.4, 0.7] by making use of two hypothetical restoration operators. One of them produced images close to the original one, and another one produced images close to the one with maximum deterioration. We found that the precision of Restoration Score was also higher than that of the SNR improvement at measurement levels outside the range [-0.4, 0.7].

5. CONCLUSIONS

In this paper, we have proposed a novel quantitative measure of image restoration quality. It has been shown by detailed experiments that the proposed measure is more precise in measuring image restoration quality, as compared with the SNR improvement. Another feature of the proposed measure is that it contains clearly defined and meaningful reference points. 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 some main limitations of the SNR improvement, and it can be a better comparison vehicle in the objective evaluation of image restoration quality.

6. REFERENCES

[1] M. I. Sezan and A. M. Tekalp, "Survey of recent developments in digital image restoration," *Optical Engineering*, vol. 29, no. 5, pp. 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] John Mandel, *Evaluation and Control of Measurements,* Marcel Dekker, Inc., New York (1991).

[5] G. L. Anderson and A. N. Netrarali, "Image restoration based on a subjective criterion," *IEEE Trans. Syst., Man, Cyber.* vol. SMC-6, no. 12, pp. 845-853, 1976.

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