Invariant error metrics for image reconstruction

J. R. Fienup

Expressions are derived for the normalized root-mean-square error of an image relative to a reference image. Different versions of the error metric are invariant to different combinations of effects, including the image's (a) being multiplied by a real or complex-valued constant, (b) having a constant added to its phase, (c) being translated, or (d) being complex conjugated and rotated 180°. Invariance to these effects is particularly important for the phase-retrieval problem. One can also estimate the parameters of those effects. Similarly, two wave fronts can be compared, allowing for arbitrary constant (piston) and linear (tilt) phase terms. One can also include a weighting function. The relation between the error metric and other quality measures is derived. © 1997 Optical Society of America

Key words: Image processing, image quality assessment, phase retrieval.

1. Introduction

When developing digital image-reconstruction algorithms, we often start with an ideal test image, degrade it in some way (say, blurring, adding noise, or discarding parts of the data), and then perform a reconstruction on the degraded image. To evaluate the performance of the algorithm, we compute a quality measure for the reconstructed image. References 1 and 2 discuss several image-quality measures. To compare the reconstructed image with the ideal, many researchers use the root-mean-square (rms) error as a measure of image quality because it is relatively easy to analyze. In some cases the reconstructed image might be multiplied by an unknown constant relative to the ideal image. If we did not want to count this as a defect, we would wish to compute a modified rms error that is invariant to multiplicative constants. When reconstructing images from Fourier magnitude data (i.e., performing phase retrieval), which is insensitive to image translation, constant phase, and complex conjugation and rotation by 180° (i.e., the twin image), we would wish to compute a modified rms error that is invariant to these effects. We wish the quality metric to be invariant to these various effects since they do not degrade the image or make it less recognizable.

Section 2 of this paper derives rms errors that are insensitive to these effects and shows how to estimate the unknown parameters. Section 3 shows how to use the same mathematics to compare two wave fronts in such a way that they are insensitive to arbitrary piston and tilt terms. The weighted version of the error metric is also given. Section 4 comments on a necessary computation—the upsampling of the cross correlation between the ideal and reconstructed images. Section 5 shows the relations among these error metrics and other error metrics. Section 6 concludes the paper.

The author is with ERIM International, Inc., P.O. Box 134008, Ann Arbor, Michigan 48113-4008.

Received 5 May 1997.
0003-6935/97/328352-06$10.00/0

2. Normalized rms Error

Let us take a given ideal digital image $f(x, y)$, defined on the integer grid, and a reconstructed image $g(x, y)$. The commonly used sum-of-squares error between them is given by

$$e^2 = \sum |g(x, y) - f(x, y)|^2, \tag{1}$$

where the summation is over all the pixels (x, y) in the image. Since this metric depends on the multiplicative scale of $f(x, y)$, it is meaningless by itself, and so we prefer to use a normalized error

$$E^2 = \frac{\sum |g(x, y) - f(x, y)|^2}{\sum |f(x, y)|^2}, \tag{2}$$

or, better yet, its square root E. This normalized sum-of-squares error metric E^2 is the normalized mean-square error (nmse) metric, and E is the normalized root-mean-square error (nrmse). Note that we allow the image to be complex valued. Next we explore versions of the nmse that are insensitive to various factors that we may deem to be unimportant.

Although in this paper we use the discrete model, which is pertinent to digital image-reconstruction experiments, everything described here is equally applicable to functions of continuous variables. Also, although we show everything in two dimensions, it is equally applicable to any number of dimensions.

An important class of problems concerns phase retrieval, for which we know the magnitude $|F(u, v)|$ of the Fourier transform of the image:

$$F(u, v) = (MN)^{-1/2} \sum f(x, y) \exp\left[-i2\pi\left(\frac{ux}{M} + \frac{vy}{N}\right)\right] = \mathcal{F}[f(x, y)], \tag{3}$$

where the summation is over $x = 0, 1, \ldots, M - 1$, and $y = 0, 1, \ldots, N - 1$, and $\mathcal{F}$ denotes the Fourier transform. For this problem

$$|\mathcal{F}[f(x, y)]| = |\mathcal{F}[\exp(i\alpha) f(x - x_0, y - y_0)]| = |\mathcal{F}[\exp(i\alpha) f^*(-x - x_0, -y - y_0)]|, \tag{4}$$

where the asterisk denotes complex conjugation, that is, the Fourier magnitude is insensitive to multiplicative constant phase factors $\exp(i\alpha)$, translations (x_0, y_0), and image twinning (complex conjugation plus rotation by 180°. The Fourier magnitude is ambiguous with respect to these alternative images. We consider any of these alternative images to be acceptable substitutes for the ideal image $f(x, y)$ since they have the same appearance as $f(x, y)$. Hence we wish our error metric to be zero for any of them. For this reason, for phase-retrieval problems we use an error metric of the form

$$E^2 = \min\left\{\min_{\alpha, x_0 y_0} E^2(g; \alpha, x_0, y_0), \min_{\alpha, x_0 y_0} E^2[g^*(-); \alpha, x_0, y_0]\right\}, \tag{5}$$

where

$$E^2(g; \alpha, x_0, y_0) = \frac{\sum |\exp(i\alpha) g(x - x_0, y - y_0) - f(x, y)|^2}{\sum |f(x, y)|^2}, \tag{6}$$

and $g^*(-)$ represents the twin image. That is, we find the parameters α, x_0, and y_0 that minimize the error of the image $g(x, y)$ with respect to $f(x, y)$, we repeat the process for the twin image, and then we pick between the image and the twin, depending on which has the lower error. Fortunately, as we show next, there is a straightforward way to compute this error metric.

To facilitate the derivation, we rewrite

$$E^2(g; \alpha, x_0, y_0) = \frac{\sum |g(x - x_0, y - y_0)|^2 + \sum |f(x, y)|^2 - \left[\exp(-i\alpha) \sum f(x, y) g^*(x - x_0, y - y_0) + \text{c.c.}\right]}{\sum |f(x, y)|^2} = \frac{\sum |g(x, y)|^2 + \sum |f(x, y)|^2 - [\exp(-i\alpha) r_{fg}(x_0, y_0) + \text{c.c.}]}{\sum |f(x, y)|^2}, \tag{7}$$

where we assume a circular (wraparound) coordinate system for $g(x, y)$ wherein no values are lost by translation, c.c. denotes the complex conjugate of the term that precedes it, and $r_{fg}(x, y)$ is the cross correlation of f and g. As is commonly known, minimizing the mean-square error is equivalent to maximizing the cross correlation.

For a given translation (x_0, y_0), we determine the constant phase α that minimizes the error by setting equal to zero the partial derivative of $E^2(g; \alpha, x_0, y_0)$ with respect to α and solving for α:

$$0 = \frac{-[-i \exp(-i\alpha) r_{fg}(x_0, y_0) + \text{c.c.}]}{\sum |f(x, y)|^2} = \frac{-2 \operatorname{Im}[\exp(-i\alpha) r_{fg}(x_0, y_0)]}{\sum |f(x, y)|^2} = \frac{2|r_{fg}(x_0, y_0)| \sin[\alpha - \arg r_{fg}(x_0, y_0)]}{\sum |f(x, y)|^2}, \tag{8}$$

which has the solution

$$\alpha = \arg r_{fg}(x_0, y_0) + n\pi, \tag{9}$$

where n is an integer and $\arg(z)$ is the phase(modulo 2π) of z. Inserting this solution for α into the expression for E^2 yields an expression with the term $\pm|r_{fg}(x_0, y_0)|$, with the plus sign corresponding to the maximum of E^2 and the minus sign corresponding to the desired minimum:

$$\min_{\alpha} E^2(g; \alpha, x_0, y_0) = \frac{\sum |g(x, y)|^2 + \sum |f(x, y)|^2 - 2|r_{fg}(x_0, y_0)|}{\sum |f(x, y)|^2}. \tag{10}$$

Noting that $r_{ff}(0, 0) = \Sigma |f(x, y)|^2$, we can express Eq. (10) in the form

$$\min_{\alpha} E^2(g; \alpha, x_0, y_0) = \frac{r_{gg}(0, 0) + r_{ff}(0, 0) - 2|r_{fg}(x_0, y_0)|}{r_{ff}(0, 0)}, \tag{11}$$

which is minimized for the location (x_0, y_0) where $|r_{fg}(x_0, y_0)|$ is maximized:

$$\min_{\alpha, x_0, y_0} E^2(g; \alpha, x_0, y_0) = \frac{r_{gg}(0, 0) + r_{ff}(0, 0) - 2 \max_{x_0, y_0} |r_{fg}(x_0, y_0)|}{r_{ff}(0, 0)}. \tag{12}$$

So, for evaluating images reconstructed for the problem of phase retrieval from Fourier magnitude data, we can compute an appropriate quality metric by (i) computing the magnitude of the cross correlation of the reconstructed image with the ideal image, (ii) finding the maximum value of its magnitude (see Section 4), (iii) evaluating Eq. (12) above, (iv) taking the square root to get the nrmse, (v) repeating steps (i)–(iv) for the twin image $g^*(-x, -y)$, and (vi) taking the minimum of the two error metrics. Note that we use $r_{ff}(0, 0)$ as a notational device only; the actual computation would be performed more efficiently as $\Sigma|f(x, y)|^2$, and similarly for $r_{gg}(0, 0)$.

In the process of computing the nrmse we have also computed (x_0, y_0), the translation that minimizes the nrmse, and we have determined whether $g(x, y)$ or its twin is closer to the ideal image, $f(x, y)$. We can also compute the phase constant a by evaluating $r_{fg}(x_0, y_0)$ and taking its phase.

The exact form of the invariant error metric changes, depending on what we assume we know or to what we want the metric to be invariant. Suppose, for example, that in the phase-retrieval problem the Fourier magnitude data is multiplied by an unknown, real, nonnegative constant. If a constant phase is unknown as well, then the reconstructed image will be multiplied by an unknown complex constant, $\alpha = \alpha_R + i\alpha_I$, where α_R is the real part and α_I is the imaginary part of α. Then we wish to minimize the error:

$$E^2(g; \alpha, x_0, y_0) = \frac{\sum|\alpha g(x - x_0, y - y_0) - f(x, y)|^2}{\sum|f(x, y)|^2} = \frac{|\alpha|^2 \sum|g(x, y)|^2 + \sum|f(x, y)|^2 - [\alpha^* r_{fg}(x_0, y_0) + \text{c.c.}]}{\sum|f(x, y)|^2}. \tag{13}$$

For a given translation (x_0, y_0), we determine the constant α_R that minimizes the error by setting equal to zero the partial derivative of $E^2(g; \alpha, x_0, y_0)$ with respect to α_R and solving for α_R:

$$\begin{aligned} 0 &= \frac{2\alpha_R \sum|g(x, y)|^2 - [r_{fg}(x_0, y_0) + \text{c.c.}]}{\sum|f(x, y)|^2} \\ &= \frac{2\alpha_R \sum|g(x, y)|^2 - 2\,\text{Re}[r_{fg}(x_0, y_0)]}{\sum|f(x, y)|^2}, \end{aligned} \tag{14}$$

which has the solution

$$\alpha_R = \frac{\text{Re}[r_{fg}(x_0, y_0)]}{\sum|g(x, y)|^2}, \tag{15}$$

and similarly,

$$\alpha_I = \frac{\text{Im}[r_{fg}(x_0, y_0)]}{\sum|g(x, y)|^2}, \tag{16}$$

yielding

$$\alpha = \frac{r_{fg}(x_0, y_0)}{\sum|g(x, y)|^2}. \tag{17}$$

Inserting α back into the error metric and simplifying yields

$$\begin{aligned} \min_{\alpha} E^2(g; \alpha, x_0, y_0) &= 1 - \frac{|r_{fg}(x_0, y_0)|^2}{\sum|g(x, y)|^2 \sum|f(x, y)|^2} \\ &= 1 - \frac{|r_{fg}(x_0, y_0)|^2}{r_{gg}(0, 0) r_{ff}(0, 0)}, \end{aligned} \tag{18}$$

which is minimized for the location (x_0, y_0), where $|r_{fg}(x_0, y_0)|$ is maximized:

$$\min_{\alpha, x_0, y_0} E^2(g; \alpha, x_0, y_0) = 1 - \frac{\max_{x_0, y_0}|r_{fg}(x_0, y_0)|^2}{r_{gg}(0, 0) r_{ff}(0, 0)}, \tag{19}$$

for the case of invariance with respect to a complex constant factor. The procedure for computing this error metric is identical to the one described above but uses the new expression for $E^2(g; \alpha, x_0, y_0)$. Note that, if we express the complex constant as $\alpha = |\alpha|\exp(ia)$ and minimize with respect to $|\alpha|$ and a rather than with respect to α_R and α_I, we get this same result. Also note the considerable difference between Eq. (19) and Eq. (12), for which invariance to only a was desired. Also note that the optimum phase constant a is the same whether or not we require invariance to $|\alpha|$.

Similarly, if we restrict the multiplicative constant to being real, assuming a known constant phase, then the error metric becomes

$$\min_{\alpha_R, x_0, y_0} E^2(g; \alpha_R, x_0, y_0) = 1 - \frac{\max_{x_0, y_0}\left(\{\text{Re}[r_{fg}(x_0, y_0)]\}^2\right)}{r_{gg}(0, 0) r_{ff}(0, 0)}. \tag{20}$$

From the definition of $E^2(g, \alpha, x_0, y_0)$ we see that $E^2(g; \alpha, x_0, y_0) \geq 0$. Furthermore, since the second term in Eq. (19) is negative, we see that

$$0 \leq \min_{\alpha, x_0, y_0} E^2(g; \alpha, x_0, y_0) \leq 1, \tag{21}$$

and similarly

$$0 \leq \min_{\alpha_R, x_0, y_0} E^2(g; \alpha_R, x_0, y_0) \leq 1. \tag{22}$$

If we wish the error metric to be invariant to only the multiplicative constant and assume that there is no translation, then in inequalities (21) and (22) we simply do not find the maximum of the cross correlation but replace $r_{fg}(x_0, y_0)$ with $r_{fg}(0, 0) = \Sigma f(x, y)g^*(x, y)$.

When making the error metric invariant to multiplication by α or α_R, we multiply $g(x, y)$ by $\alpha = r_{fg}{}^*(x_0, y_0)/\Sigma|g(x, y)|^2$ or by $\alpha_R = \mathrm{Re}[r_{fg}(x_0, y_0)]/\Sigma|g(x, y)|^2$. Note that these terms differ in magnitude from the normalization constant

$$\mathrm{const} = \left[\frac{\sum|f(x, y)|^2}{\sum|g(x, y)|^2}\right]^{1/2}, \tag{23}$$

which would cause the energy of const $\times$ $g(x, y)$ to equal the energy of $f(x, y)$. Multiplication by α_R ordinarily yields an error metric that is smaller than that obtained by multiplication of g by const, and multiplication by α yields the smallest error metric of all for complex-valued images.

If we wish the error metric to be invariant to only the translation and assume that there is no multiplicative constant, then we compute

$$\begin{aligned}\min_{x_0,y_0} E^2(g; x_0, y_0) &= \frac{\sum|g(x - x_0, y - y_0) - f(x, y)|^2}{\sum|f(x, y)|^2}\\ &= \frac{\sum|g(x, y)|^2 + \sum|f(x, y)|^2 - 2 \max_{x_o, y_o} \mathrm{Re}[r_{fg}(x_0, y_0)]}{\sum|f(x, y)|^2}\\ &= \frac{r_{gg}(0, 0) + r_{ff}(0, 0) - 2 \max_{x_o, y_o} \mathrm{Re}[r_{fg}(x_0, y_0)]}{r_{ff}(0, 0)}.\end{aligned} \tag{24}$$

3. Fourier Domain and Weighting

Since $\Sigma f(x, y)g^*(x, y) = \Sigma F(u, v)G^*(u, v)$ by Parseval's theorem, we have

$$\begin{aligned}E^2(g; \alpha, x_0, y_0) &= \frac{\sum|\alpha g(x - x_0, y - y_0) - f(x, y)|^2}{\sum|f(x, y)|^2}\\ &= \frac{\sum|\alpha G(u, v)\exp[-i2\pi(ux_0/M + vy_0/N)] - F(u, v)|^2}{\sum|F(u, v)|^2}.\end{aligned} \tag{25}$$

Similar relations hold if we replace α with α_R or $\exp(ia)$. All the preceding derivations can take place in the Fourier domain as well as in the image domain.

Suppose we wish to compare an estimated wave front $G(u, v)$ with an ideal wave front given by $F(u, v)$, but we want the error metric to be invariant to piston and tilt wave-front errors. With α replaced by $\exp(ia)$ in Eq. (25), we have exactly that error metric. Hence for calculating the error of a wave front or for least-squares matching of wave fronts we can use the same computations. With this notation the tilt term in u has the coefficient $-2\pi x_0/\mathrm{M}$, which causes a spatial-domain translation by x_0, and similarly for the v term. The values of a, x_0, and y_0 found by these steps give the least-squares solution to the piston and tilt phases of $G(u, v)$ relative to $F(u, v)$.

If we are matching wave fronts and know the way in which the signal-to-noise ratio (SNR) varies across the wave front, then we can introduce a weighting function to optimize the matching. Similarly, in an image-reconstruction problem we may wish to have a Fourier domain weighting function that emphasizes certain spatial frequencies or eliminates the consequences of bad (u, v)-plane pixels. If we multiply both $F(u, v)$ and $G(u, v)$ by a weighting function $W(u, v)$, we obtain the weighted error metric

$$\begin{aligned}E^2(g; \alpha, x_0, y_0) &= \frac{\sum|\alpha W(u, v)G(u, v)\exp[-i2\pi(ux_0/M + vy_0/N)] - W(u, v)F(u, v)|^2}{\sum|W(u, v)F(u, v)|^2}\\ &= \frac{\sum|\alpha g_w(x - x_0, y - y_0) - f_w(x, y)|^2}{\sum|f_w(x, y)|^2},\end{aligned} \tag{26}$$

where

$$f_w(x, y) = \mathscr{F}^{-1}[W(u, v)F(u, v)] = f(x, y) * w(x, y), \quad (27)$$

where the asterisk denotes convolution and $w(x, y) = \mathscr{F}^{-1}[W(u, v)]$. Then we perform the same calculations as above but replace $f(x, y)$ with $f_w(x, y)$, and $g(x, y)$ with $g_w(x, y)$. The weighting function can also be, for example, that which is due to aperture weighting (apodization) used to control the sidelobes of the impulse response or one that reflects the transfer function of the human visual system.

4. Upsampling the Cross Correlation

Unless we know the image translation to be an integer pixel distance (or zero), it is important to upsample the cross-correlation function $r_{fg}(x, y)$ to determine accurately the peak of its magnitude or of its real part. This is most simply accomplished by performance of the cross correlation by Fourier techniques:

$$r_{fg}(x, y) = \mathscr{F}^{-1}[G^*(u, v)F(u, v)], \quad (28)$$

where, prior to the inverse Fourier transform, we embed $[G^*(u, v)F(u, v)]$ in a large array of zeros. Then we can find the true maximum magnitude of the cross correlation with good accuracy by selecting the maximum from the upsampled pixels. This approach is suitable for small arrays but not for arrays so large that the embedded array becomes difficult to keep in memory or difficult to fast Fourier transform.

One approach investigated for computing the upsampled maximum of $r_{fg}(x, y)$ for larger arrays was to (i) compute $r_{fg}(x, y)$ without upsampling and find the maximum of $r_{fg}(x, y)$ to within the nearest pixel, (ii) extract a moderate-sized array about the maximum of $r_{fg}(x, y)$ and upsample that to the desired oversampling ratio, and (iii) find the maximum of the upsampled array. This approach sometimes fails because of ringing artifacts associated with the edges of the extracted array. This ringing can be greatly reduced by the application of a function to weight down the edges of the extracted array prior to upsampling.

Another approach to consider is again to compute the maximum of $r_{fg}(x, y)$ to within the nearest pixel; then, from that starting point we perform a nonlinear optimization to determine the maximum of $r_{fg}(x, y)$ to within a small fraction of a pixel. A nonlinear optimization algorithm, such as a conjugate-gradient search or a Newton method, is aided by the existence of analytic expressions for the derivatives of $r_{fg}(x, y)$ with respect to x and y, and these can be computed easily.

We have found that, for some cases, the cross-correlation peak may be highly asymmetric about its maximum, making attempts to estimate the peak location by fitting of a simple parabola to the peak to be of limited utility in those cases. Fitting a higher-order two-dimensional polynomial would be required in such cases.

5. Relations to Other Metrics

First we relate the invariant nmse to the standard deviation of the phase error and the Strehl ratio, then to the SNR. In Ref. 3 it was shown that, if the phase errors dominate over the magnitude errors, then for a Gaussian-distributed phase error with a standard deviation σ_ϕ (in radians) the expected nmse (with shifts ignored) is given by

$$\begin{aligned}\langle E^2\rangle &= \left\langle \min_\alpha E^2(g; \alpha)\right\rangle \\ &= \min_\alpha \frac{\left\langle \sum |\alpha g(x, y) - f(x, y)|^2\right\rangle}{\sum |f(x, y)|^2} \\ &= 1 - \exp(-\sigma_\phi{}^2), \end{aligned} \quad (29)$$

which approaches unity for $\sigma_\phi{}^2 \gg 1$ and approaches $\sigma_\phi{}^2$ when $\sigma_\phi{}^2 \ll 1$. Angle brackets denote the expected value over the Gaussian distribution.

For the phase error $\phi(u, v)$ over an aperture $A(u, v)$, the (coherent) impulse response is $c_\phi(x, y) = \mathscr{F}^{-1}\{A(u, v)\exp[i\phi(u, v)]\}$ and the point-spread function is $s_\phi(x, y) = |c_\phi(x, y)|^2$. The Strehl ratio is given by[4]

$$I_S = \frac{s_\phi(0, 0)}{s_0(0, 0)} \approx 1 - \sigma_\phi{}^2 \approx \exp(-\sigma_\phi{}^2) \quad (30)$$

for small phase errors, where $s_0(0, 0)$ is the peak of the impulse response when $\phi = 0$. We see that

$$\langle E^2\rangle \approx 1 - I_S, \qquad I_S \approx 1 - \langle E^2\rangle \quad (31)$$

when Fourier phase errors dominate over magnitude errors. These relations are accurate whenever the Fourier magnitude is fairly uniformly distributed, which is the case for coherent (complex-valued, speckled) images and for incoherent (real, nonnegative) images that are collections of points. For low-contrast, extended, incoherent images, most of the signal is concentrated near the origin in Fourier space, and the error metric tends to be smaller than what is predicted by approximations (31) for a given value of I_S or $\sigma_\phi{}^2$.

If we think of the integrated squared noise as the numerator for our error metric and the integrated squared signal as the denominator, then we can consider E^2 to be the reciprocal of the (power) SNR. This does not hold, however, when the SNR is near to or less than unity since the factor α substantially decreases the effect of the noise in that regime. The normalization factor α prevents the nrmse from becoming greater than unity, even when $1/\text{SNR} > 1$. Assuming noise that is uncorrelated with the signal, we can show that

$$\langle E^2\rangle \approx \frac{1}{\text{SNR} + 1}, \qquad \text{SNR} \approx \frac{1 - \langle E^2\rangle}{\langle E^2\rangle}. \quad (32)$$

6. Conclusions

We have shown that we can make the nmse between an ideal image and a reconstructed image invariant to the effects of multiplicative constants, constant phases, translations, and image twinning (complex conjugation and rotation by 180°). Calculating the invariant error metrics requires finding the peak of the cross correlation of the ideal and reconstructed images, and we must do this to subpixel accuracy. We can similarly compute an error metric for evaluating the difference between two wave fronts that is invariant to piston and tilt terms. Fourier domain weighting functions are easily included. The normalized invariant squared error metric is approximately equal to one minus the Strehl ratio and to the inverse of one plus the (power) SNR.

Portions of this paper were presented at the 1996 Annual Meeting of the Optical Society of America.

References

1. E. H. Linfoot, *Fourier Methods in Optical Image Evaluation* (The Focal Press, London, 1964).
2. A. Eskicioglu and P. S. Fisher, "Image quality measures and their performance," IEEE Trans. Commun. **12,** 2959–2965 (1995).
3. J. R. Fienup and A. M. Kowalczyk, "Phase retrieval for a complex-valued object by using a low-resolution image," J. Opt. Soc. Am. A **7,** 450–458 (1990).
4. M. Born and E. Wolf, *Progress in Optics* (MacMillan, New York, 1964), Sec. 9.1.3.