A Visual Model Weighted Cosine Transform for Image Compression and Quality Assessment

NORMAN B. NILL

Abstract—**Utilizing a cosine transform in image compression has several recognized performance benefits, resulting in the ability to attain large compression ratios with small quality loss. Also, incorporation of a model of the human visual system into an image compression or quality assessment technique intuitively should (and has often proven to) improve performance. Clearly, then, it should prove highly beneficial to combine the image cosine transform with a visual model. In the past, combining these two has been hindered by a fundamental problem resulting from the scene alteration that is necessary for proper cosine transform utilization. A new analytical solution to this problem, taking the form of a straightforward multiplicative weighting function, is developed in this paper. This solution is readily applicable to image compression and quality assessment in conjunction with a visual model and the image cosine transform. In the development, relevant aspects of a human visual system model are discussed, and a refined version of the mean square error quality assessment measure is given which should increase this measure's utility.**

I. INTRODUCTION

NUMEROUS image compression techniques have been developed over the years which have as their goal the reduction of the number of bits needed to transmit or store digital imagery—consistent with implementation complexity and cost constraints. The chosen technique must, of course, maintain acceptable image quality upon final image decompression, reconstruction, and display. Several review papers on this subject have been published in recent years; the ones by Netravali and Limb [1] and Jain [2] are particularly comprehensive. It is generally recognized that one of the most efficient and effective methods for obtaining large compression ratios with manageable complexity is transform compression and, in partciular, cosine transform compression [3].

The cosine transform has been found to perform better overall than other transforms in many digital image compression applications [3]. The cosine transform greatly reduces the blocking problem—related to the undesirable Fourier-Gibbs phenomenon or "ringing" at edges—when small subimage pixel blocks (typically 16 × 16 pixels) are individually transformed, encoded, transmitted, decoded, and then added to form the reconstructed image. (Applying transforms to subimages rather than the entire image as a whole has advantages in terms of allowing adaptivity to differences in scene parts—see Habibi [4].) In addition, it has been shown that the cosine transform is a limiting case of the Karhunen-Loeve transform [5], [6]. The latter is an optimum transform for image compression in the mean square sense, but requires knowledge of the source correlation for implementation—information which is seldom available—whereas the cosine transform is a deterministic transform not requiring such knowledge. Finally, the all-real cosine transform can be rapidly computed by applying fast discrete transform algorithms [21].

Now, in conjunction with the proven utility of transform techniques applied to imagery compression, one could intuitively expect that if a suitable model of the *human visual system (HVS)* could be successfully combined with the compression process, an improvement in compression performance would result. This expectation follows from the fact that, in most compression applications, a human is the final observer of the imagery operated upon. Application of various models of the HVS have in fact been empirically found to improve (noncosine) transform compression performance [7]-[11].

Along the same lines of reasoning, incorporation of an HVS model in an image quality measure should result in a quantitative correlate of human quality assessment. Developing such a measure leads to greater insight, efficiency, and exactness in the process of developing a prospective compression scheme since it should entail rapid, precise, mathematically tractable, digital computations in place of time-consuming, labor-intensive, subjective (human) quality assessments [11]-[13].

The preceding discussion indicates that 1) a combination of the cosine transform together with an appropriate HVS model would lead to a digital image compression scheme exhibiting improved performance; and 2) the performance of a compression scheme would be accurately quantifiable if an objective image quality measure which also incorporates an HVS model is utilized. There is, however, a fundamental problem in combining an HVS model with an image cosine transform which stems from the input scene alteration necessary for proper cosine transform utilization. This problem has neither been given adequate attention in the literature nor has it been adequately solved. This paper describes this problem and its ramifications, then develops a mathematically and physically meaningful and useful solution.

Section II presents the fundamental problem associated with combining an image cosine transform with an HVS model. Section III then describes the selection of an appropriate HVS model and its incorporation in an objective image quality measure, followed by a discussion in Section IV of how to incorporate the HVS model in transform image compression. With this basic methodology of use established, Section V solves the HVS/cosine transform problem in as mathematically rigorous a fashion as is possible, while also being consistent with practical considerations. Section VI looks at the implications of this solution.

II. PROBLEM STATEMENT

In order to combine the cosine transform of imagery with an HVS model, the two must be both mathematically and physically compatible for application to either compression or quality assessment. We assume that an adequate mathematical compatibility is achieved if we can correctly combine the cosine transform with the HVS model in a linear systems theoretic sense. Given the current imperfect state of knowledge of the HVS and the fact that, in many applications,

Paper approved by the Editor for Signal Processing and Communication Electronics of the IEEE Communications Society for publication without oral presentation. Manuscript received August 16, 1984; revised February 11, 1985. This work was supported by the Rome Air Development Center, Griffiss AFB, NY, under Contract F19628-83-C-0001, as MITRE Mission-Oriented Investigation and Experimentation Project 7330.

The author is with the MITRE Corporation, Bedford, MA 01730.

0090-6778/85/0600-0551$01.00 © 1985 IEEE

little *a priori* knowledge of scenes or viewing conditions is available, it is appropriate to model the HVS as a stationary linear system. This first-order approximation to the complex process of vision has been successfully applied in the past to image processing problems [8]–[10].

We thereby arrive at the basic model of interest: a previously processed scene is input to a linear system (the HVS), is modified by that system's impulse response function, and is displayed or recorded at the output (i.e., at the brain). The basic problem in using the cosine transform in this model (in the scene processing stage) revolves around the fact that the processing of the scene must include a basic scene alteration in order to correctly apply the cosine transform to it. This necessary alteration takes the form of forcing a symmetry onto a normally asymmetrical original scene, even though it is a likeness of the asymmetrical scene that will be viewed by the human observer—*not* the altered (symmetrical) scene. It is this necessary forced symmetry which makes it difficult to rigorously combine the desired linear systems theory with the physical cases of interest. This important property of the cosine transform, when used in image processing, has not been adequately brought out in the literature and is therefore explored in more depth in the following.

It is well known that (see, for example, Bracewell [14]): 1) if a function $g(x)$ is even, i.e., $g(x) = f(-x) + f(x)$ where $f(-x) = f(x)$, and 2) if $g(x)$ is all real, then the Fourier transform of $g(x)$ reduces exactly to the cosine transform of $g(x)$, which in turn equals the cosine transform of $f(x)$; i.e.,

$$\int_{-\infty}^{\infty} g(x)e^{-2\pi j v x}\,dx \equiv \int_{-\infty}^{\infty} g(x)\cos(2\pi v x)\,dx$$

$$= 2\int_{0}^{\infty} f(x)\cos(2\pi v x)\,dx. \qquad (1)$$

Although this relation has been known for many years in a purely mathematical sense, its practical application and utility in signal data compression has been identified only fairly recently by Ahmed, Natarajan, and Rao [15].

Now let $f(x)$ represent the intensity distribution of a one-dimensional (1-D), asymmetrical scene[1] as a function of distance x. Forcing a scene to be symmetrical, as represented by the formation of $g(x)$, then allows application of the cosine transform in place of the Fourier transform with no loss of information. That is, the scene can be exactly reconstructed from just the cosine transform.

Now, from stationary linear systems theory it is known that the convolution integral and Fourier convolution theorem hold, as given in (2) and (3), respectively:

$$\int_{-\infty}^{\infty} f(t)h(x-t)\,dt = \text{output} \qquad (2)$$

where

$f(x)$ = input scene

$h(x)$ = HVS impulse response function

$$\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} f(t)h(x-t)\,dt\right] e^{-2\pi j v x}\,dx = F(v)H(v) \qquad (3)$$

where $F(v)$, $H(v)$ are the Fourier transforms of $f(x)$, $h(x)$, respectively.

[1] In most of this paper, a 1-D continuous analysis is performed for ease of understanding of the concepts and derivations involved. The conversion to 2-D is performed where necessary.

Equations (2) and (3) are highly useful results in the context of application to imagery transform compression and quality assessment, and we would like to utilize these relations with the cosine transform. As was stated, however, in applying the cosine transform, the original scene $f(x)$, which is asymmetrical (noneven), must first be made symmetrical (even). This is easily accomplished by forming $g(x)$ from $f(x)$ as previously discussed. Since $g(x)$ is even and real (real because light intensity is all real), the Fourier transform of $g(x)$ equals the cosine transform of $g(x)$ as in (1). Furthermore, since $h(x)$ is real due to physical constraints on the visual process and even by assumption, the convolution theorem in (2) therefore holds for the cosine transform as well as for the Fourier transform. That is, because $g(x)$ and $h(x)$ are real, even functions,

$$\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} g(t)h(x-t)\,dt\right]\cos(2\pi v x)\,dx$$

$$= F_c(v)H_c(v) = F_c(v)H(v) \qquad (4)$$

where $F_c(v)$, $H_c(v)$ are the cosine transforms of $f(x)$, $h(x)$, respectively.

Using a different line of reasoning, Clarke [34] has demonstrated that the cosine transform of an input signal, corresponding to a given Fourier spatial frequency and arbitrary phase, is only significant in a small neighborhood of that Fourier spatial frequency. Clarke concluded that it is feasible to utilize a multiplicative factor [in this case $H_c(v)$] in conjunction with the cosine transform of a signal [in this case $F_c(v)$]. Even though (4) is true in a mathematical sense, however, it is not sufficient by itself in a physically meaningful sense for application to image processing. In other words, a contradiction results from the fact that in order to utilize the cosine transform, the scene must be altered, but this very alteration causes the loss of a necessary physical significance, since the human observer is not viewing this altered scene. This contradiction is the fundamental problem that must be overcome in order to properly combine an image cosine transform with an HVS model. It will be seen that (4) can be used as a starting point in the problem solution.

Before discussing solution approaches, however, we first look at just how the cosine transform, in combination with the HVS model, could be used in image quality assessment and image compression.

III. HVS Model and Image Quality Assessment

The image quality measure, actually a measure of quality degradation, that has most often been used in digital image compression research is the mean square error (MSE) between the original, unprocessed image and the processed image. However, it has often been empirically determined that the MSE and its variants do not correlate well with subjective (human) quality assessments [9], [10], [13]. The reasons are not well understood, but one suspects that the MSE does not adequately track the types of degradations caused by digital image compression processing techniques, and that it does not adequately "mimic" what the human visual system does in assessing image quality.

Some researchers [16], [17] have attempted to improve upon quality assessment by incorporating elaborate models of the visual process. Such models have been devised in an attempt to simulate the effects of many of the parameters affecting vision, such as orientation, field angle, and Mach bands, but their utility for practical problems is small due to their complexity, inherent unknowns, and need for sometimes detailed *a priori* knowledge of viewing condition parameter values. Incorporation of an elaborate visual system model into an image quality measure is not practical at present.

However, it has been found that several simplifying assumptions for the visual model can still lead to a quality measure that performs better than, for instance, the MSE, which does not incorporate a visual model [10], [13]. If one assumes that the visual system is linear, at least for low contrast images, and is isotropic, and that the scenes viewed are monochrome and static, with observer-preferred luminance levels, and are viewed for an observer-preferred length of time, then these assumptions lead to a single, straightforward function representing the visual system, which is amenable to incorporation in a quality measure.[2] These assumptions are valid for certain classes of image observation, notably reconnaissance images being viewed for interpretation purposes.

More specifically, many researchers have measured the human threshold contrast sensitivity to periodic patterns (sine waves, square waves, etc.) viewed at a range of spatial frequencies–good reviews of this work can be found in Levi [18] and Kelly [19]. By taking the reciprocal of such a contrast sensitivity curve, one arrives at a curve akin to the spatial frequency response function of the visual system. Mathematically, applying linear systems concepts, this is equivalent to the Fourier transform of the response of the visual system to an impulse light stimulus. Fig. 2 shows the shape of this visual spatial frequency response function $H(r)$.

Even after incorporating a visual response function in a quality measure, however, a further refinement is in order to more closely mimic how a human assesses quality. In many classes of image observation, a good assumption is that the observer will base his/her judgment of overall scene quality on the higher structural (activity) regions contained in the scene. Thus, an improvement to an overall scene quality measure should become apparent by incorporating a weighting factor that puts more emphasis on high structure subimage areas and less emphasis on low structure subimage areas. A concept similar to this postulate has in fact been found to improve quality assessment performance [13].

Bringing together the preceding concepts of a visual response function and subimage structure weighting, and working within the framework of the MSE difference between the original and processed images as well as within the framework of stationary linear systems theory, a quality measure results that, it is felt, will more accurately track human assessments of quality. This quality measure can be stated in the 2-D discrete Fourier (for now) spatial frequency domain as given in (5):

$$K^{-1}\sum_{i=1}^{B} W_i \sum_{u=0}^{M-1}\sum_{v=0}^{N-1} H^2(r)[F_i(u,v)-\hat{F}_i(u,v)]^2 \tag{5}$$

where

- B = number of subimage blocks in scene
- K = normalization factor such as total energy
- $H(r)$ = rotationally symmetric spatial frequency response of HVS, $r = \sqrt{u^2+v^2}$
- $F_i, \hat{F}_i$ = Fourier transform of unprocessed and processed subimage i, respectively
- M, N = number of Fourier coefficients +1, in orthogonal u, v directions
- W_i = subimage i structure weighting factor, proportional to subimage's intensity level variance (see Chen and Smith [20])
- W_i = 1.0 for maximum structure subimage
- W_i $\rightarrow$ 0.0, for minimum structure subimage.

[2] An initial nonlinearity is sometimes introduced into an HVS model by preprocessing the image with a logarithmic or power function [12]. Here we are particularly interested in low contrast images, however, whereby we can assume to be working in a linear region of a (possible) overall nonlinearity [32]. In any case, it is not at all clear what nonlinear function would be most appropriate to use [33], and none is used in this study.

IV. HVS Model in Image Compression

After the cosine transform of a subimage is obtained, a variety of encoding schemes can be applied to produce compression of the cosine coefficients. One promising technique is to reorder the 2-D array of subimage cosine coefficients into a more tractable 1-D array, as suggested by Tescher [22]. This conversion to 1-D readily invites application of various simple yet effective schemes to produce compression. Note that some refinement to the 1-D "zig-zag" coefficient reordering developed by Tescher can be accomplished if it is replaced with 1-D radial coefficient reordering (see (13) and surrounding discussion). In radial reordering, constant radial frequencies are grouped together to achieve, on the average, a monotonically decreasing 1-D curve.

As an example of use, a 1-D least squares polynomial could be fit to the reordered 1-D cosine coefficients, whereupon the polynomial coefficients, rather than the cosine coefficients, would be encoded. If a polynomial order sufficiently less than the number of cosine coefficients can adequately reconstruct those cosine coefficients (upon decompression), then good compression has been achieved. In this approach, the HVS model is incorporated by weighting the residual squared error between a given cosine coefficient and the prospective polynomial fit value, according to the relative weight the eye–brain system would give to that particular spatial frequency, as exemplified by the HVS model in the spatial frequency domain (see Fig. 2). In this way, the accuracy of the polynomial fit at a given spatial frequency is directly proportional to the relative emphasis placed on that spatial frequency by the HVS. This approach also avoids the necessity of taking out the HVS model in the decoding/decompression steps.

Other, nonpolynomial schemes for compressing/encoding the 1-D reordered cosine coefficients are possible[3] and would incorporate the HVS model in essentially the same manner as described in the preceding example. This application, of course, presupposes a solution to the HVS/cosine problem, which is given in Section V.

V. HVS/Cosine Problem Solution

The problem is to correctly combine a linear systems model of the HVS with the cosine transform of imagery, for the image quality assessment and image compression applications described in Sections III and IV. The problem can be avoided by treating the HVS model and image cosine transform separately, as suggested by Hall [9]. This procedure, however, although correct, is computationally very intensive. As shown in Fig. 1(a), it requires four separate 2-D complex Fourier transforms of the image in addition to the two 2-D cosine transforms that actually form the compression/decompression scheme.

Griswold [24] developed a solution to the problem in the power spectrum domain whereby a circular symmetric bit map was sought and obtained. The development follows a set of predetermined rules regarding the form of the solution function, borrowed from properties of Fourier power spectra and autocorrelation functions. However, in this author's opinion, these particular solution function construction rules do not have a bearing on the form of a solution in the cosine transform domain. In essence, a solution function has been force-fitted to conform to certain preconditions that cannot be adequately justified for their applicability to the problem at hand.

[3] Research in image compression/encoding approaches based on 2-D to 1-D cosine transform coefficient reordering is currently active at The MITRE Corporation; see Sullivan *et al.* [23].

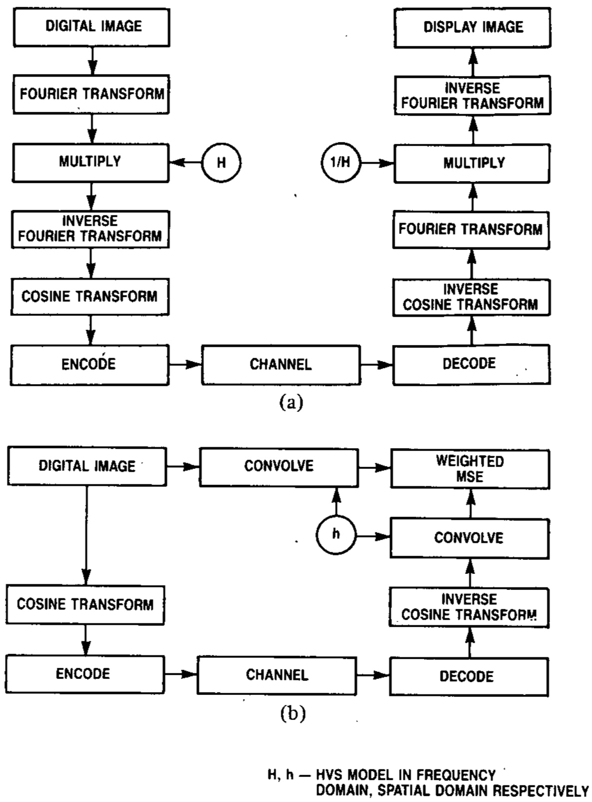

Fig. 1. Mathematically correct but computationally intensive and, therefore, generally impractical methods for combining the human visual system (HVS) model with the image cosine transform for (a) image transform compression and (b) image quality assessment.

As for the quality measure given in (5), one could apply the equivalent equation in the spatial (image) domain, thus avoiding the HVS/cosine problem, as shown in Fig. 1(b). This approach, however, necessitates the application of two 2-D convolutions, a generally computationally intensive procedure. In addition, it would not directly aid in identifying the visual spatial frequency range of importance for compression optimization in the cosine transform domain.

In short, a generally viable solution has not been reported in the literature, to this author's knowledge. What is sought is a fairly simply utilized function that can be directly applied to the HVS model in the spatial frequency domain and to the image cosine transform, such that combining the two becomes both a theoretically correct procedure and a practically useful one.

Referring back to (4), it was seen that the HVS model $h(x)$ was convolved with the altered, symmetrical scene $g(x)$, whereas the convolution of $h(x)$ with $f(x)$ is really desired—since $f(x)$ represents the unaltered scene. The latter can be achieved by bringing in a unit step function $u(x)$ as follows:

$$\text{given: } h(x) \otimes g(x) \tag{6}$$

where $\otimes$ denotes convolution

$$\text{then: } h(x) \otimes [u(x)g(x)] = h(x) \otimes f(x) \tag{7}$$

$$\text{where } u(x) = \begin{cases} 1, & x \geqslant 0 \\ 0, & x < 0 \end{cases}$$

Converting the left-hand side of (7) to the Fourier transform spatial frequency domain results in

$$H(v)\left[\left(\frac{1}{2}\,\delta(v) - \frac{j}{2\pi v}\right) \otimes F_c(v)\right]. \tag{8}$$

Using the convolution property associated with the Dirac delta function, $\delta(v)$, and rearranging terms of (8) yields

$$\frac{1}{2}\,H(v)F_c(v) - jH(v)\left[\frac{1}{2\pi v} \otimes F_c(v)\right]. \tag{9}$$

The quantity enclosed in brackets in (9) is seen to be the Hilbert transform of $F_c(v)$, which equals the sine transform $F_s(v)$, of $f(x)$. Thus, (9) contains the cosine and sine transforms of $f(x)$, which together completely specify the Fourier transform of $f(x)$. By computing $F_s(v)$ from $F_c(v)$, the full Fourier transform of $f(x)$ would be generated, which would solve the problem of how to combine $H(v)$ in a linear systems theoretic sense when starting with $F_c(v)$. However, this approach would also defeat the purpose of using the cosine transform in the first place (see Section I). Instead, we will accept the basic condition that only the image cosine transform is to be computed. But is there a reasonably straightforward function which, when multiplied by the image cosine transform and HVS model, results in a good approximation to (9)? That is, we seek a function $A(v)$ such that

$$A(v)H(v)F_c(v) = \frac{1}{2}\,H(v)F_c(v) - jH(v)\left[\frac{1}{2\pi v} \otimes F_c(v)\right]. \tag{10}$$

Rearranging terms in (10) and dividing out $H(v)$,

$$A(v) = \frac{1}{2} - \frac{j[v^{-1} \otimes F_c(v)]}{2\pi F_c(v)}. \tag{11}$$

Now for tractability $j = \sqrt{-1}$ must be eliminated in (11), which can be accomplished[4] by taking the modulus of $A(v)$,

$$|A(v)| = \frac{1}{2}\left[1 + \frac{[v^{-1} \otimes F_c(v)]^2}{\pi^2 F_c^{\,2}(v)}\right]^{1/2}. \tag{12}$$

In order to be able to apply the same $|A(v)|$ to every image cosine transform computed, a value for $F_c(v)$ in (12) is needed that is representative of all scenes to be encountered. It is reasonable in this regard to consider the individual scenes as single realizations of a stochastic process. Indeed, a scene can be accurately modeled statistically as a random step process [26]–[28] which results in a negative exponential autocorrelation function and a 1-D power spectrum, $P(v)$, taking the form

$$P(v) = \frac{s}{(2\pi v)^2 + \alpha^2} + \bar{I}^2\delta(v) \tag{13}$$

where s is related to the scene intensity variance, α equals the reciprocal of the average intensity pulse width, and $\bar{I}^2$ is the scene's average intensity level. If the scenes are assumed to be wide-sense stationary, then the statistics (joint probability distribution) of the scene $f(x)$ will be identical to the statistics of the altered scene $g(x) = f(x) + f(-x)$. Hence, the power spectrum (as calculated from the scene statistics) representing the scene equals the power spectrum representing the altered scene, which leads to

$$|\tilde{F}_c(v)|^2 = P(v) \tag{14}$$

[4] In optics, for example, the modulus alone often proves to adequately represent the full Fourier transform in linear systems spatial frequency applications—see, for example, Smith [25]. In this application it is also compatible with $H(v)$ and $F_c(v)$.

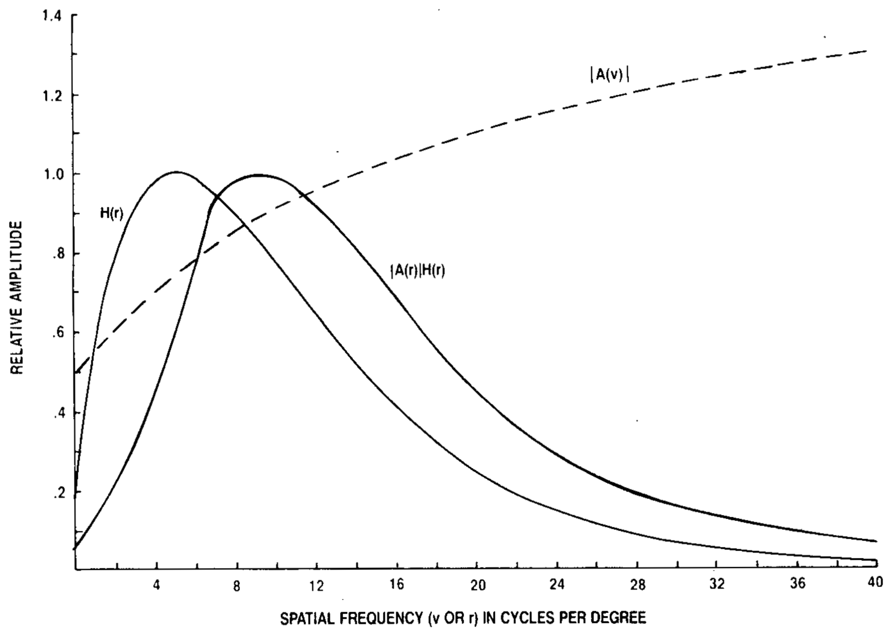

Fig. 2. $H(r)$: the normalized 2-D spatial frequency response of the human visual system; $|A(v)|$: the 1-D correction factor which allows use of $H(r)$ with an image cosine transform; $|A(r)|H(r)$: the 2-D (normalized) function which replaces $H(r)$ in image cosine (as opposed to Fourier) transform applications. The mathematical forms of these curves are given in (19), (18), and (20), respectively.

where the tilde implies the cosine transform of the representative average scene. For convenience, the scene intensity bias level $\bar{I}^2$ can be set equal to zero, which is equivalent to simply rescaling the scene intensity values, and results in [combining (13) and (14)]

$$|\tilde{F}_c(v)|^2 = \frac{s}{(2\pi v)^2 + \alpha^2}. \tag{15}$$

Then substituting (15) into (12) and working with the positive square root of (15) yields

$$|A(v)| = \frac{1}{2}\left[1 + \left(\frac{1}{\pi^2}\right)\frac{\left[v^{-1} \otimes \sqrt{\dfrac{s}{(2\pi v)^2 + \alpha^2}}\right]^2}{\dfrac{s}{(2\pi v)^2 + \alpha^2}}\right]^{1/2}. \tag{16}$$

The numerator in (16) is evaluated in closed form in Appendix A with the result

$$v^{-1} \otimes \sqrt{\frac{s}{(2\pi v)^2 + \alpha^2}} = 2\sqrt{\frac{s}{(2\pi v)^2 + \alpha^2}}\, \log_e\left[\frac{2\pi}{\alpha} v + \sqrt{\frac{4\pi^2}{\alpha^2} v^2 + 1}\right]. \tag{17}$$

Finally, substituting (17) into (16) gives the solution function in closed form

$$|A(v)| = \left[\frac{1}{4} + \frac{1}{\pi^2}\left[\log_e\left(\frac{2\pi}{\alpha} v + \sqrt{\frac{4\pi^2}{\alpha^2} v^2 + 1}\right)\right]^2\right]^{1/2} \tag{18}$$

The 1-D $|A(v)|$ function (18) is shown in Fig. 2 for $\alpha = 11.636$ degrees^{-1}, where this typical value of the parameter α is derived in Appendix B. An appropriate rotationally symmetric HVS model is also shown in Fig. 2 for

$$H(r) = (0.2 + 0.45r)e^{-0.18r} \tag{19}$$

where the radial frequency r is in cycles per degree of visual angle subtended. This particular curve is based on the work of Mannos and Sakrison [10] and DePalma and Lowry [29]. The former identified an $H(r)$ with a peak at 8 cycles/degree by a well-conceived experimental trial and error process, albeit utilizing digital imagery raised to a power as the input. The latter arrived at $H(r)$ (actually its inverse) by psychophysical spatial frequency threshold measurements on human observers, resulting in an $H(r)$ peak around 3 cycles/degree. This author has found the $H(r)$ curve based on DePalma and Lowry's results to perform quite well in image quality work done a few years ago (unpublished) on nondigital imagery. Considering all of these factors, it is felt that the $H(r)$ in Fig. 2, which is a composite derived from [10] and [29], is a good working representation of the HVS.

Now, since the general image is 2-D, it is more appropriate to utilize the 2-D version of $|A(v)|$. It would, however, be very difficult if not impossible to obtain this 2-D version directly from (18) in closed form. Instead, this conversion can be conveniently carried out numerically (for given α) by utilizing any one of several methods, as discussed by Marchand [30]. The steps in the numerical procedure used here were: 1) compute the Abel transform of $H(r)$ to convert $H(r)$ to 1-D, i.e., to $H(v)$; 2) multiply $|A(v)|$ by $H(v)$; 3) compute the 1-D cosine transform of the even function $|A(v)|H(v)$; and 4) compute the 1-D Hankel transform of the result of step 3. This 2-D rotationally symmetric function (i.e., the result of step 4) is shown in Fig. 2. A curve fit to this numerically obtained function is

$$|A(r)|H(r) = \begin{cases} 0.05e^{r^{0.554}}, & \text{for } r < 7 \\ e^{-9[|\log_{10} r - \log_{10} 9|]^{2.3}}, & \text{for } r \geq 7. \end{cases} \tag{20}$$

This function can now be treated in image cosine transform applications in the same manner as $H(r)$ would be treated in image Fourier transform applications. For example, for quality assessment one simply substitutes $|A(r)|H(r)$ from Fig. 2 for the $H(r)$ in (5), when dealing with image cosine transforms instead of image Fourier transforms.

VI. Discussion and Conclusions

By utilizing the multiplicative weighting factor $|A(v)|$ derived in this paper, together with a human visual system model $H(r)$, it becomes correct in a physical sense as well as a linear systems sense (to within the assumptions made in Section V) to combine an image cosine transform with a human visual system model. It is believed that this combining of an HVS model with the image cosine transform will result in better performance in image compression and image quality assessment applications. In quality assessment, performance should also be enhanced by inclusion of the subimage structure weighting given in (5).

As can be seen in Fig. 2, the effect of $|A(v)|$ is to translate $H(r)$ into the more positive frequency direction, resulting in a higher frequency peak, relatively less emphasis on the lower (than peak) frequencies, and relatively more emphasis on the higher (than peak) frequencies. This implies that the higher spatial frequencies in the cosine transform domain play a more important role in the corresponding human-observed image quality than they do in the "equivalent" Fourier transform domain (where $A(v) = 1.0$ in the latter). Thus, encoding of image cosine transform coefficients for bit compression requires more attention and emphasis on maintaining the fidelity of these higher frequencies than would be needed for image Fourier transform coefficients. The actual curve for $|A(r)|H(r)$ is a function of the original $H(r)$ and the several parameters noted in Appendix B affecting $|A(v)|$. With the realistic values utilized here, $H(r)$ and $|A(r)|H(r)$ have peak values at 5.2 and 9.0 cycles/degree, respectively. Other parameter values and $H(r)$ curves could, of course, result in different peak locations, but the trend and implications for all reasonable values remain as stated above.

Appendix A

Evaluation of

$$v^{-1} \otimes \sqrt{\frac{s}{(2\pi v)^2 + \alpha^2}}$$

for (16): It is more convenient to evaluate this convolution integral by utilizing the Fourier convolution theorem (3), thereby considering the two components separately:

$$\int_{-\infty}^{\infty} v^{-1} e^{2\pi j v x}\, dv = \int_{-\infty}^{\infty} \cancelto{0}{\frac{\cos(2\pi v x)}{v}}\, dv + j\int_{-\infty}^{\infty} \frac{\sin(2\pi v x)}{v}\, dv = j\pi \operatorname{sgn}(x) \tag{A-1}$$

where

$$\operatorname{sgn}(x) = \begin{cases} 1, & x \geq 0 \\ -1, & x < 0 \end{cases}$$

$$\sqrt{s}\int_{-\infty}^{\infty} \frac{e^{2\pi j v x}}{\sqrt{4\pi^2 v^2 + \alpha^2}}\, dv = \sqrt{s}\int_{-\infty}^{\infty} \frac{\cos(2\pi v x)}{\sqrt{4\pi^2 v^2 + \alpha^2}}\, dv + j\sqrt{s}\int_{-\infty}^{\infty} \cancelto{0}{\frac{\sin(2\pi v x)}{\sqrt{4\pi^2 v^2 + \alpha^2}}}\, dv \tag{A-2}$$

$$= \frac{\sqrt{s}}{\pi} K_0(\alpha|x|) \tag{A-3}$$

where $K_0(\cdot)$ is the modified Bessel function of the second kind, of order zero.

Then combining (A-1) and (A-2) and taking the inverse Fourier transform,

$$v^{-1} \otimes \sqrt{\frac{s}{(2\pi v)^2 + \alpha^2}}$$

$$= \int_{-\infty}^{\infty} \left[j\pi \operatorname{sgn}(x) \frac{\sqrt{s}}{\pi} K_0(\alpha|x|) \right] e^{-2\pi j v x}\, dx \tag{A-4}$$

$$= j\sqrt{s}(-2j) \int_0^{\infty} K_0(\alpha|x|) \sin(2\pi v x)\, dx \tag{A-5}$$

(since $\operatorname{sgn}(x)$ is odd and $K_0(\alpha|x|)$ is even)

$$= 2\sqrt{\frac{s}{(2\pi v)^2 + \alpha^2}} \log_e \left[\frac{2\pi}{\alpha} v + \sqrt{\frac{4\pi^2}{\alpha^2} v^2 + 1} \right] \tag{A-6}$$

where integral tables can be used to obtain (A-3) and (A-6) (see [31, eq. 333.78a and 541.8c]).

Appendix B

The scene can be modeled as a pulse process resulting in the scene power spectrum given in (13), where α is equal to the reciprocal of the average pulse width. With v in cycles/degree, it is appropriate to have α in units of reciprocal degrees, which can be arrived at by calculating the angle subtended by the average pulse width of the displayed scene from the human observer's location. From geometry this angle is equal to

$$\alpha^{-1} = 2 \arcsin \left[\frac{W}{2\sqrt{\frac{W^2}{4} + D^2}} \right]$$

where

W = average pulse width (in linear units)
D = observer to displayed-scene distance.

For a digitized scene that has been quantized to about 8 bits of grey levels and is displayed on a CRT, W is on the order of 1.5 pixels, where a typical CRT display pixel is effectively 0.5 mm on a side; then, with a typical value for D of 0.5 m, this results in

$$\alpha = 11.636 \text{ (degrees)}^{-1}.$$

REFERENCES

[1] A. N. Natravali and J. O. Limb, "Picture coding: A review," *Proc. IEEE,* vol. 68, pp. 366–406, Mar. 1980.

[2] A. K. Jain, "Image data compression: A review," *Proc. IEEE,* vol. 69, pp. 349-389, Mar. 1981.

[3] A. G. Tescher, "Transform image coding," in *Advances in Electronics and Electron Physics,* suppl. 12, W. K. Pratt, Ed. New York: Academic, 1979, ch. 4, pp. 113-155.

[4] A. Habibi, "Survey of adaptive image coding techniques," *IEEE Trans. Commun.,* vol. COM-25, pp. 1275-1284, Nov. 1977.

[5] N. Ahmed and M. D. Flickner, "A derivation for the discrete cosine transform," *Proc. IEEE,* vol. 70, pp. 1132-1134, Sept. 1982.

[6] K. S. Shanmugan, "Comments on discrete cosine transform," *IEEE Trans. Comput.,* vol. C-24, p. 759, July 1975.

[7] J. O. Limb, "Visual perception applied to the encoding of pictures," in *Advances in Image Transmission Techniques, Proc. Soc. Photo-Opt. Instrum. Eng.,* vol. 87, pp. 80-87, 1976.

[8] T. G. Stockham, "Image processing in the context of a visual model," *Proc. IEEE,* vol. 60, pp. 828-842, July 1972.

[9] C. F. Hall, "Digital color image compression in a perceptual space," Ph.D. dissertation, Dep. Elec. Eng., Univ. Southern California, Los Angeles, Jan. 1978.

[10] J. L. Mannos and D. J. Sakrison, "The effects of a visual fidelity criterion on the encoding of images," *IEEE Trans. Inform. Theory,* vol. IT-20, pp. 525-536, July 1974.

[11] D. J. Granrath, "The role of human visual models in image processing," *Proc. IEEE,* vol. 69, pp. 552-561, May 1981.

[12] C. F. Hall and E. L. Hall, "A nonlinear model for the spatial characteristics of the human visual system," *IEEE Trans. Syst., Man, Cybern.,* vol. SCM-7, pp. 161-170, Mar. 1977.

[13] F. X. J. Lukas and Z. L. Budrikis, "Picture quality prediction based on a visual model," *IEEE Trans. Commun.,* vol. COM-30, pp. 1679-1692, July 1982.

[14] R. N. Bracewell, *The Fourier Transform and Its Applications.* New York: McGraw-Hill, 1965.

[15] N. Ahmed, T. Natarajan, and K. R. Rao, "Discrete cosine transform," *IEEE Trans. Comput.,* vol. C-23, pp. 90-93, Jan. 1974.

[16] I. Overington, "Toward a complete model of photopic visual threshold performance," *Opt. Eng.,* vol. 21, pp. 2-13, Jan.-Feb. 1982.

[17] A. Schnitzler, "Effects of spatial frequency filtering on the performance of the composite photographic-human visual system," *Photogr. Sci. Eng.,* vol. 21, pp. 209-215, July-Aug. 1977.

[18] L. Levi, "Vision in communication," in *Progress in Optics,* vol. 8, E. Wolf, Ed. New York: Elsevier, 1970, sect. 7, pp. 345-374.

[19] D. H. Kelly, "Visual contrast sensitivity," *Opt. Acta,* vol. 24, pp. 107-130, Feb. 1977.

[20] W. Chen and C. H. Smith, "Adaptive coding of monochrome and color images," *IEEE Trans. Commun.,* vol. COM-25, pp. 1285-1292, Nov. 1977.

[21] W. Chen, C. H. Smith, and S. Fralick, "A fast computational algorithm for the discrete cosine transform," *IEEE Trans. Commun.,* vol. COM-25, pp. 1004-1009, Sept. 1977.

[22] A. G. Tescher, "A dual transform coding algorithm," in *Proc. Nat. Telecommun. Conf.,* Nov. 1981, pp. C9.2.1-C9.2.3.

[23] D. R. Sullivan, N. B. Nill, R. D. Braun, D. H. Lehman, and B. W. Fam, "Advanced imagery lab final report," and D. R. Sullivan, B. W. Fam, R. D. Braun, and N. B. Nill, "Advanced imagery laboratory image compression research," RADC Tech. Reps., to be published, 1985.

[24] N. C. Griswold, "Perceptual coding in the cosine transform domain," *Opt. Eng.,* vol. 19, pp. 306-311, May-June 1980.

[25] W. J. Smith, *Modern Optical Engineering.* New York: McGraw-Hill, 1966, ch. 11, pp. 308-324.

[26] L. E. Franks, "A model for the random video process," *Bell Syst. Tech. J.,* vol. 45, pp. 609-630, Apr. 1966.

[27] Y. Itakura, S. Tsutsumi, and T. Takagi, "Statistical properties of the background noise for the atmospheric windows in the intermediate infrared region," *Infrared Phys.,* vol. 14, pp. 17-29, 1974.

[28] D. Halford, "A general mechanical model for $|f|^\alpha$ spectral density random noise with special reference to flicker noise $1/|f|$," *Proc. IEEE,* vol. 56, pp. 251-258, Mar. 1968.

[29] J. J. DePalma and E. M. Lowry, "Sine wave response of the visual system. II. Sine wave and square wave contrast sensitivity," *J. Opt. Soc. Amer.,* vol. 52, pp. 328-335, Mar. 1962.

[30] E. W. Marchand, "Derivation of the point spread function from the line spread function," *J. Opt. Soc. Amer.,* vol. 54, pp. 915-919, July 1964.

[31] W. Gröbner and N. Hofreiter, *Integraltafel-Zweiter Teil-Bestimmte Integrale.* New York: Springer-Verlag, 1966.

[32] T. N. Cornsweet, *Visual Perception.* New York: Academic, 1970, p. 334.

[33] C. N. Nelson, "The theory of tone reproduction," in *The Theory of the Photographic Process,* C. E. K. Mees and T. H. James, Eds. New York: MacMillan, 1966, ch. 22, pp. 470-477.

[34] R. J. Clarke, "Spectral response of the discrete cosine and Walsh-Hadamard transforms," *IEE Proc.,* vol. 130, part F, pp. 309-313, June 1983.

★

Norman B. Nill received the B.S. degree in photographic science from Rochester Institute of Technology, Rochester, NY, performed initial graduate studies in optics at the University of Rochester, Rochester, and received the M.S. degree in electrical engineering from New York University, New York, NY.

Over a ten-year period at Perkin-Elmer, Danbury, CT, starting in 1968, he was involved in analytical and experimental work in image science, including image evaluation, coherent optical image processing, and work on the NASA Space Telescope. From 1978 to 1981 he worked in acoustic signal processing at Analysis and Technology, North Stonington, CT, and image pattern recognition at Synectics, Fairfax, VA. Since he joined the MITRE Corporation, Bedford, MA, in 1981, he has been involved in concepts development and assessment of advanced techniques for image processing and exploitation. He is the author of several articles in the image science area.

Mr. Nill is a member of the Society of Photographic Scientists and Engineers.