Structural similarity quality metrics in a coding context: exploring the space of realistic distortions

Alan C. Brooks, *Member, IEEE*, Xiaonan Zhao, *Student Member, IEEE*, and
Thrasyvoulos N. Pappas, *Fellow, IEEE*

Abstract— Perceptual image quality metrics have explicitly accounted for human visual system (HVS) sensitivity to subband noise by estimating just noticeable distortion (JND) thresholds. A recently proposed class of quality metrics, known as structural similarity metrics (SSIM), models perception implicitly by taking into account the fact that the HVS is adapted for extracting structural information from images. We evaluate SSIM metrics and compare their performance to traditional approaches in the context of realistic distortions that arise from compression and error concealment in video compression/transmission applications. In order to better explore this space of distortions, we propose models for simulating typical distortions encountered in such applications. We compare specific SSIM implementations both in the image space and the wavelet domain; these include the complex wavelet SSIM (CWSSIM), a translation-insensitive SSIM implementation. We also propose a perceptually weighted multi-scale variant of CWSSIM, which introduces a viewing distance dependence and provides a natural way to unify the structural similarity approach with the traditional JND-based perceptual approaches.

Index Terms— Error concealment, human perception, image quality, structural similarity, video coding, video compression.

I. INTRODUCTION

MOST existing objective fidelity metrics compare the reference and distorted images on a point-by-point basis, whether this is done in the original image domain, as in mean squared error based metrics such as peak signal to noise ratio (PSNR), or in a transform domain, such as the perceptually weighted subband/wavelet or discrete cosine transform (DCT) domain [1]. The most advanced of these metrics are based on low-level models of the HVS. On the other hand, a recently proposed class of quality metrics, known as Structural SIMilarity (SSIM) [2], accounts for high-level HVS characteristics and allows substantial point-by-point distortions that are not perceptible, such as spatial and intensity shifts, as well as contrast and scale changes. Our goal is to evaluate SSIM metrics and to compare their performance to traditional approaches in the context of realistic distortions

A. C. Brooks is with the Defensive Systems Division, Northrop Grumman Corporation, Rolling Meadows, IL 60008 USA (e-mail: alancbrooks@gmail.com).

X. Zhao and T. N. Pappas are with the Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL 60208 USA (e-mail: xiaonan-zhao@northwestern.edu; pappas@eecs.northwestern.edu).

that arise from compression and error concealment in video compression/transmission applications. In order to explore this space of distortions in an efficient and systematic manner, we propose models for simulating typical distortions encountered in such applications.

Perceptual image quality metrics have relied on explicit low-level models of human perception that account for sensitivity to subband noise as a function of frequency, local luminance, and contrast/texture masking [1], [3]. Typically, the signal is analyzed into components (e.g., spatial and/or temporal subbands), and the role of the perceptual model is to provide the maximum amount of distortion that can be introduced to each component without resulting in any perceived distortion. This is usually referred to as the *just noticeable distortion* level or *JND*. While these metrics were developed for near-threshold applications, they have also been used in suprathreshold applications [4], [5]. The main idea is to normalize the distortion by the JND [6], [7]. More systematic studies of the suprathreshold case have been conducted by Hemami *et al.* [8]–[11]. However, while perceptual metrics can successfully account for subband (frequency) dependence of the HVS sensitivity to noise and contrast and luminance masking, they cannot account for imperceptible structural changes, such as spatial shifts, intensity shifts, contrast changes, and scale changes.

The SSIM metrics [2] are based on high-level properties of the HVS, but employ no explicit model of the HVS. They are derived from assumptions about the high-level functionality of the HVS, and in particular, account for the fact that it is adapted for extracting structural information (relative spatial covariance) from images. Thus, they can more effectively quantify suprathreshold compression artifacts, as such artifacts tend to distort the structure of an image. Even though the SSIM metrics are not based on explicit models or measurements of HVS sensitivities, they implicitly account for important HVS properties such as light adaption and masking, in addition to the perception of image structure [2]. However, while the SSIM metrics have been shown to have a number of desirable properties, they have not been systematically studied in the context of image and video compression artifacts.

In an increasing number of applications, such as video transmission over bandlimited and noisy channels, there is a need to achieve very high compression ratios. In such cases, a certain amount of perceived distortion is unavoidable. Thus, there is an increased need for quantitative objective measures of perceived distortion. In this paper, we examine the performance of SSIM metrics for such suprathreshold video transmission applications and compare their performance to

traditional approaches.

In typical video transmission applications one encounters a variety of distortions due to source coding (quantization) and packet losses, which are concealed by different techniques. In order to isolate the different types of distortion, analyze their severity (as perceived by the HVS), and evaluate how well they correspond to metric predictions, we propose models for typical distortion artifacts such as DCT coefficient quantization, spatial interpolation concealment, temporal replacement concealment, and DC coefficient loss. These models can be used to generate typical video coding and concealment artifacts using still images, which considerably simplifies the computational cost for the simulations, and provides flexibility in isolating and controlling the severity of specific types of distortions. We demonstrate that these models make it easy to explore the space of realistic distortions, and provide an invaluable tool for metric development and evaluation.

We compare specific SSIM index implementations both in the image space and the wavelet domain. These include the complex wavelet SSIM (CWSSIM), a translation-insensitive SSIM implementation. We also propose a perceptually weighted multi-scale variant of the complex wavelet SSIM (WCWSSIM). We show that the perceptual weighting is necessary for distortions that are critically dependent on viewing distance, such as white noise and DCT-based compression. Moreover, it provides a natural way to unify the structural similarity approach with traditional JND-based perceptual approaches. Our experimental results indicate that, while structural similarity metrics – and especially the CWSSIM and WCWSSIM – provide significant advantages over traditional approaches, they also have important limitations. We believe that the proposed distortion models will be the key to addressing such limitations.

The paper is organized as follows. In Section II, we review traditional perceptual metrics. In Section III, we review the motivation, development, and theory behind structural metrics, and discuss two specific implementations, SSIM and CWSSIM. Section IV describes the results of using structural metrics to assess image quality of a variety of suprathreshold distortions. In Section V, we propose models of source coding and error concealment distortions in video compression, and evaluate SSIM techniques; we also extend CWSSIM to account for viewing distance, and show that there is a natural way to unify the SSIM approach with the JND-based perceptual approaches.

II. REVIEW OF PERCEPTUAL METRICS WITH EXPLICIT VISUAL MODELS

In this section we review perceptual metrics that have been developed for image compression applications. Such metrics incorporate explicit models of the HVS that account for the spatial contrast sensitivity function, contrast masking, and luminance masking [1], [3]. They are typically based on a multiscale frequency decomposition such as a discrete wavelet transform (DWT), subband decomposition, or discrete cosine transform (DCT). For each subband coefficient, a noise visibility threshold $t(i,k)$ is obtained that represents the amount of

noise that can be added to the coefficient without resulting in any perceived distortion. This is the *just noticeable distortion* level or *JND*. Here k denotes the subband index and i the coefficient location in the (possibly downsampled) subband image. The image distortion is then computed as follows:

$$D_p = \frac{1}{N} \left(\sum_{i,k} \max \left\{ \frac{\left| \hat{b}(i,k) - b(i,k) \right|}{t(i,k)}, 1 \right\}^{Q_s} \right)^{\frac{1}{Q_s}} \quad (1)$$

where $b(i,k)$ is the subband coefficient of the reference image, $\hat{b}(i,k)$ is corresponding coefficient of the distorted image, and $t(i,k)$ is the visibility threshold. The value of Q_s can be based on psychophysical experiments. In this paper, we will use $Q_s = 2$, as in [6]; for a discussion of other choices, see [1] and references therein. Note that when the difference of two coefficients is below the visibility threshold, the noise is invisible no matter what the actual value of the difference is; otherwise, the difference is normalized by the JND. The threshold $t(i,k)$ is the product of three terms

$$t(i,k) = t_b(k) \, \tau_l(i,k) \, \tau_t(i,k) \quad (2)$$

where $t_b(k)$ is the baseline sensitivity threshold, $\tau_l(i,k)$ is the luminance masking adjustment, and $\tau_c(i,k)$ is the contrast masking adjustment. If we combine the two masking adjustments into $\tau_m(i,k)$ and assume that we are in a suprathreshold regime, the image distortion can be expressed as follows:

$$D_p^2 = \frac{1}{N} \sum_k \frac{1}{\{t_b(k)\}^2} \sum_i \frac{1}{\{\tau_m(i,k)\}^2} \left| \hat{b}(i,k) - b(i,k) \right|^2 \quad (3)$$
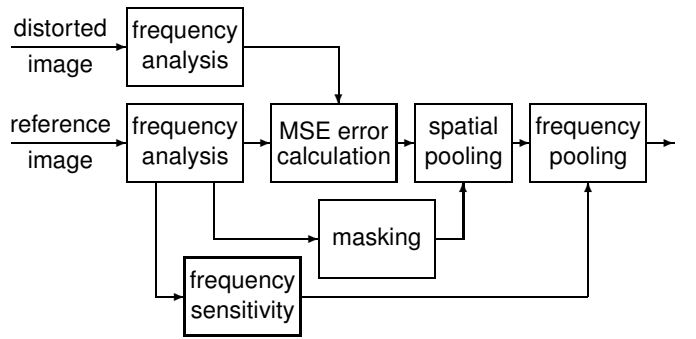
This is illustrated in Fig. 1.

Fig. 1.   Perceptual quality metric

In order to be consistent with the traditional PSNR (where $\text{PSNR} = 10 log_{10}(255^2/\text{MSE})$), we express the perceptual metric in terms of "visual decibels (dB)" [1], [12]. We define the "masked PSNR" or "perceptual PSNR (P-PSNR)" as

$$\text{P-PSNR} = 10 \log_{10} \frac{255^2}{D_p^2}. \quad (4)$$

Note that the maximum value P-PSNR is 48.13 dB, which results when all the terms in (1) are equal to 1, and hence $D_p = 1$.

In this paper, we will use the metric developed by Safranek and Johnston for subband coders [6], which we found to

have the best performance among other similar metrics [5]. A detailed description of the other metrics can be found in [1]. We will assume a fixed viewing distance of six image heights. Even though the Safranek-Johnston metric was developed for near-threshold applications, it can also be used in supra-threshold applications [4], [5]. More systematic studies of the suprathreshold case can be found in [8]–[11].

III. STRUCTURAL APPROACH TO IMAGE QUALITY MEASUREMENT

A. SSIM Review

The motivation behind the structural similarity approach for measuring image quality is that the HVS is not designed for detecting imperfections and "errors" in images. Instead, the HVS has evolved so that it can do visual pattern recognition in order to be able to extract the *structure* or connectedness of natural images. Based on this observation, it makes sense that a useful perceptual quality metric would emphasize the structure of scenes over the lighting effects. The idea that image quality metrics can be created on the basis of this philosophy was first explored in [13] and then modified, implemented, evaluated, and developed in [2]. The structural similarity approach is mostly *insensitive* to the distortions that lighting changes create: changes in the mean and contrast of an image. On the other hand, the structural approach is *sensitive* to distortions that break down natural spatial correlation of an image, such as blur, block compression artifacts, and noise.

As described in [14], the structural philosophy can be implemented using a set of equations defining the SSIM quality metric in image space. Luminance, contrast, and structure are measured separately. Given two images (or image patches) $\mathbf{x}$ and $\mathbf{y}$ to be compared, *luminance* is estimated as the mean of each image

$$\mu_x = \frac{1}{N} \sum_{n=1}^{N} x_n, \tag{5}$$

contrast is estimated using standard deviation as

$$\sigma_x = \sqrt{\frac{1}{N-1} \sum_{n=1}^{N} (x_n - \mu_x)^2}, \tag{6}$$

and *structure* is estimated from the image vector $\mathbf{x}$ by removing the mean and normalizing by the standard deviation

$$\varsigma_x = \frac{\mathbf{x} - \mu_x}{\sigma_x}. \tag{7}$$

Then, the measurements $\mu_x, \mu_y, \sigma_x, \sigma_y, \varsigma_x, \varsigma_y$ are combined using a luminance comparison function $l(\mathbf{x}, \mathbf{y})$, a contrast comparison function $c(\mathbf{x}, \mathbf{y})$, and a structure comparison function $s(\mathbf{x}, \mathbf{y})$ to give a composite measure of structural similarity:

$$SSIM(\mathbf{x}, \mathbf{y}) = l(\mathbf{x}, \mathbf{y})^\alpha \cdot c(\mathbf{x}, \mathbf{y})^\beta \cdot s(\mathbf{x}, \mathbf{y})^\gamma, \tag{8}$$

where α, β, γ are positive constants used to weight each comparison function.

The comparison functions are given as:

$$l(\mathbf{x}, \mathbf{y}) = \frac{2\mu_x \mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \tag{9}$$

$$c(\mathbf{x}, \mathbf{y}) = \frac{2\sigma_x \sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \tag{10}$$

$$s(\mathbf{x}, \mathbf{y}) = \frac{\langle \varsigma_x, \varsigma_y \rangle + C_3}{\sigma_x \sigma_y + C_3} = \frac{\sigma_{xy} + C_3}{\sigma_x \sigma_y + C_3} \tag{11}$$

where $\langle \rangle$ is the inner-product operator defining the correlation between the structure of the two images.

In this paper, we follow the example in [2] setting $\alpha = \beta = \gamma = 1$ and $C_3 = C_2/2$ to get the specific SSIM quality metric

$$SSIM(\mathbf{x}, \mathbf{y}) = \frac{(2\mu_x \mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}. \tag{12}$$

B. Complex Wavelet SSIM Review

As suggested in [15], it is straightforward to implement a structural similarity metric in the complex wavelet domain. As more wavelet-based image and video coding techniques are coming into use, it makes sense to be able to implement image quality metrics in this domain. In addition, if an application requires an image quality metric that is unresponsive to spatial translation, this extension of SSIM can be adapted so that it has low sensitivity to small translations. This requires an overcomplete wavelet transform such as the steerable pyramid [16], for which *phase information is available*. Fig. 2 shows a example of such a decomposition with three scales and four orientations. Note the octave spacing of the radial subbands.

Given complex wavelet coefficients $\mathbf{c}_x$ and $\mathbf{c}_y$ that correspond to image patches $\mathbf{x}$ and $\mathbf{y}$ that are being compared, the complex wavelet structural similarity (CWSSIM) is given by:

$$CWSSIM(\mathbf{c}_x, \mathbf{c}_y) = \frac{2|\sum_{i=1}^{N} c_{x,i} c_{y,i}^*| + K}{\sum_{i=1}^{N} |c_{x,i}|^2 + \sum_{i=1}^{N} |c_{y,i}|^2 + K}, \tag{13}$$

where K is a small positive constant set to 0.03 in this paper. This equation differs from (12) because the wavelet filters we use (with the exception of the baseband) are bandpass (i.e., they have no response at zero frequency), thus forcing the mean of the wavelet coefficients to zero ($\mu_x = \mu_y = 0$). Note also that in (13) we have substituted the expanded expressions for σ_{xy}, σ_x, and σ_y.

The overall similarity of two images is estimated as the average of the CWSSIM metric values over all spatial locations and subbands (or a subset of the the subbands, e.g., one radial band and all angular bands). In the following experiments, we compute the CWSSIM metric using a two-scale, 16-orientation steerable filter decomposition (as in [15]), and average over all the subbands except the high-pass ones (as opposed to [15], which uses the second scale averaged over all orientations.)

Wang and Simoncelli note that the wavelet coefficient phase is the key factor that determines the results of CWSSIM: "the structural information of local image features is mainly contained in the relative phase patterns of the wavelet coefficients" [15]. Linear and uniform phase changes correspond to lighting (brightness and contrast) distortions to which CWSSIM is not sensitive because the *structure* is not perturbed. Phase changes that vary irregularly from one coefficient to the next produce structural distortion and therefore low CWSSIM.
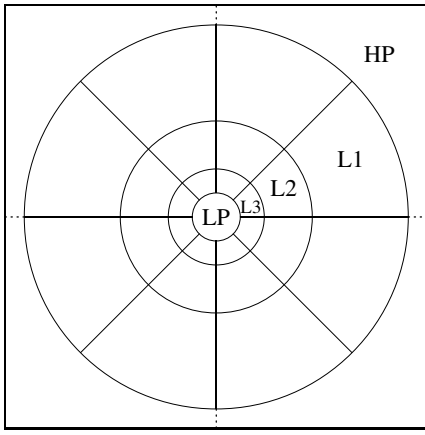
Fig. 2.   Steerable Filter Decomposition (3 scales, 4 orientations).

IV. USING SSIM TO ASSESS IMAGE QUALITY

A. Suprathreshold Effectiveness

Structural similarity can distinguish between structural and non-structural distortions, giving results that agree with perception visibly distorted images (suprathreshold distortions). The structural similarity metrics take values from in the range from 0.0 to 1.0, where zero corresponds to a loss of all structural similarity and one corresponds to having an exact copy of the original image.[1]

An example of the effectiveness of structural similarity in measuring suprathreshold distortion is depicted in Fig. 3. The original image is shown in the upper left, then five distorted images with equal PSNR, including JPEG compression [17], blur, Gaussian white noise, mean intensity shift, and contrast stretch distortions. Note that the perceived quality varies substantially from the quantized and blurred images (worst), to the noise image, to the mean and contrast modifications (best). Both structural similarity metrics, SSIM and CWSSIM, compute image quality measurements that correspond well with the idea that change in lighting (mean intensity shift, contrast stretch) has little impact on image quality, while changes that affect local relationships between pixels severely degrade image quality. The last three images (rotation, zoom, and spatial shift) have lower PSNR but the perceived quality is about as high as that of the other two images in the bottom row. The SSIM fails to predict the quality of these three images because of its sensitivity to translations, while the CWSSIM does quite well, as it has low sensitivity to small translations. The figure also shows the WCWSSIM, which will be discussed in Section V-D. Note that small amounts of scaling and rotation can be locally approximated as small translations [15]. In addition, the CWSSIM does a much better job at predicting the relative quality of the noise, blur, and JPEG images than the SSIM, because it is implemented in the subband domain, and considers distortions at different scales.

The P-PSNR predicts the quality of the first four images quite well,[2] does worse with the contrast stretch, and poorly with small translations. Overall, perceptual image quality

metrics that are based on HVS sensitivity to just-noticeable distortions are not expected to provide meaningful measurements at suprathreshold levels of distortion [5], even though the Safranek-Johnston metric does relatively well in this case. Note also, that even though the perceptual metrics developed in [8]–[11] have been specifically adapted to suprathreshold distortions, they are not designed and are not expected to be insensitive to translation, scaling, and rotation. SSIM-based image quality metrics, on the other hand, have a better chance of dealing effectively with suprathreshold distortions because they are focusing on the top-down image formation concept that the local structure of images is the most important aspect of image quality.

B. Effect of Window Size

When using SSIM metrics to compare the quality between two images, it is useful to calculate the local distortion between corresponding image patches at many locations. This allows the metric to adapt to the local statistical characteristics at different image locations. The individual quality measurements can then be combined to give a single number that represents the similarity between the images. Applications that need to measure image quality with minimum computation may only want to compute the metric at a few locations within the image [18]. However, for this paper, we measure the similarity with a sliding window at every pixel location, giving a SSIM distortion map. As suggested in [2], we use a circular Gaussian weighting function on the image patches being compared to smooth the similarity map, and we combine the measurements using a mean operator.

The choice of the $W \times W$ window size provides a balance between SSIM's ability to adapt to local image statistics and its ability to accurately compute the statistics within an image patch. A large window allows accurate statistical estimation, at the cost of being less sensitive to fine image distortions. For typical 512×512 images, a window size within the range of 7×7 to 15×15 offers a reasonable operating region.

The effect of window size on SSIM is illustrated in Figs. 4 and 5. The first image in the top row of Fig. 5 depicts the "Lena" image distorted with white Gaussian noise, while the first image in the bottom row shows "Lena" with JPEG compression noise; in both cases, the PSNR is 28.5 dB. The other images show the SSIM similarity map for $W \times W$ windows of sizes W equal to 3, 7, 15, and 31, respectively. The dark regions represent the highest SSIM distortion. Note that a $W = 3$ window is too small to accurately compute the distortions, giving an almost randomly distributed distortion map (especially in the bottom row), while a $W = 31$ window is too large to adapt to local statistics, resulting in a blurry error map and image quality scores that are too high in most regions. The distortion maps computed with $W = 7$ more accurately correspond to human perception. Of course, the window size should depend on viewing distance, an issue to which we return in Section V-D. Note that white Gaussian noise is most noticeable in the smooth (low frequency) regions of the image

---

[1]Note that the image domain SSIM implementation can also take negative values when the local image structure is inverted.

[2]Note that the image mean is typically subtracted before the metric calculation.

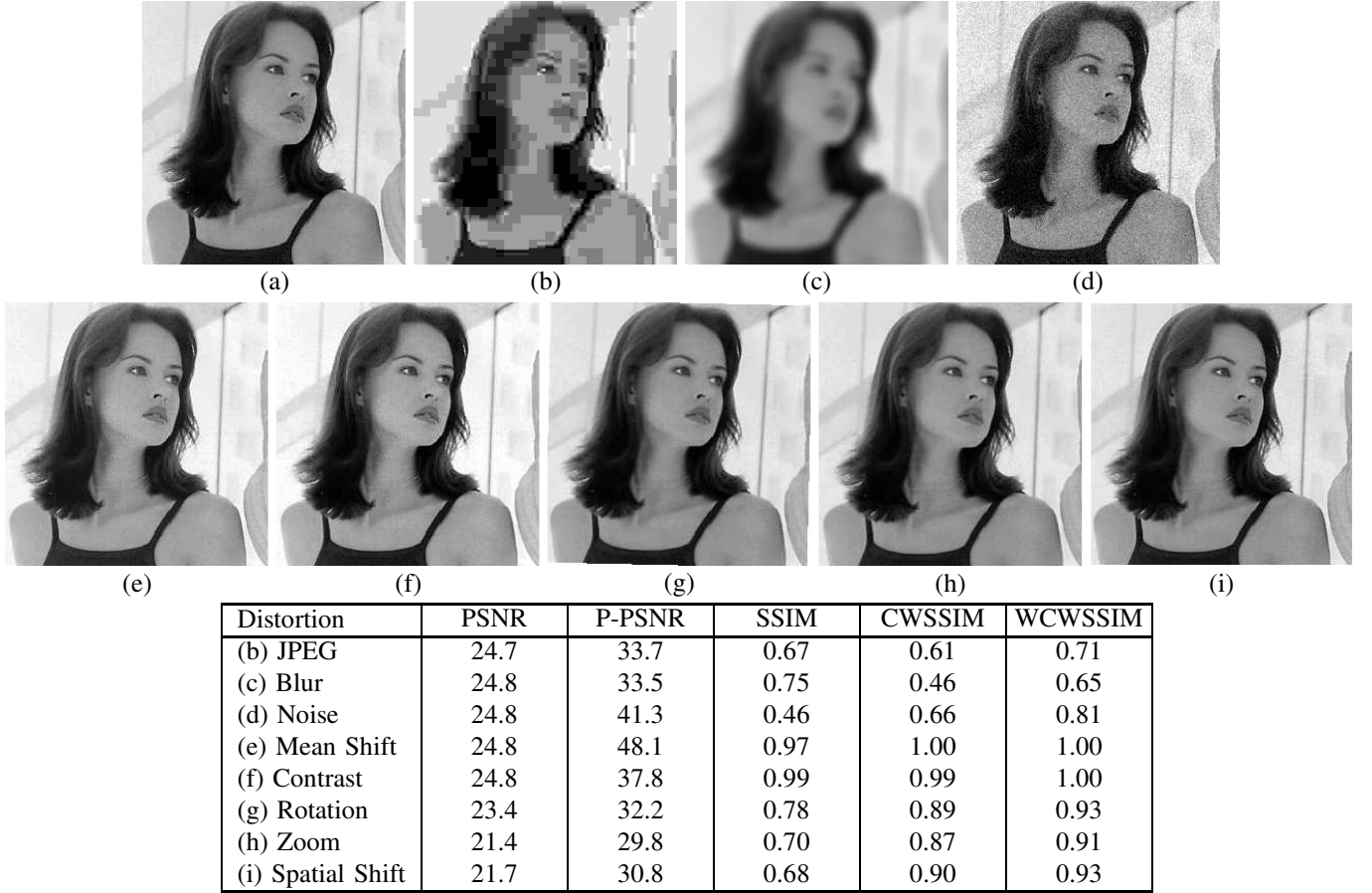| Distortion	PSNR	P-PSNR	SSIM	CWSSIM	WCWSSIM
(b) JPEG	24.7	33.7	0.67	0.61	0.71
(c) Blur	24.8	33.5	0.75	0.46	0.65
(d) Noise	24.8	41.3	0.46	0.66	0.81
(e) Mean Shift	24.8	48.1	0.97	1.00	1.00
(f) Contrast	24.8	37.8	0.99	0.99	1.00
(g) Rotation	23.4	32.2	0.78	0.89	0.93
(h) Zoom	21.4	29.8	0.70	0.87	0.91
(i) Spatial Shift	21.7	30.8	0.68	0.90	0.93

Fig. 3. The effect of different distortions on PSNR, SSIM, CWSSIM, weighted CWSSIM, and perceptual PSNR metrics. (a) Original; (b) JPEG compression; (c) blur; (d) white noise; (e) mean shift; (f) contrast change; (g) rotation (1.3^o); (h) zoom in; (h) spatial shift (right by 2 pixels).

such as the shoulder and background. This demonstrates the strong inherent masking effect of the SSIM metric, which is primarily due to the contrast comparison term $c(\mathbf{x}, \mathbf{y})$ in (8). (When the variance of the original image is high, then $c(\mathbf{x}, \mathbf{y})$ is high, i.e., the noise is masked.) In the JPEG compression example, the dark regions correspond well with the most perceptually annoying compression artifacts: the blockiness on the shoulder, cheeks, and background. Note that the masking effect is not as strong, because to some extend, JPEG exploits masking for compression. (If it did a good job, the distortion map would be relatively uniform.)

The relationship between SSIM and window size is also shown in Fig. 4, which plots the overall quality pooled over the image for each window size. This plot is derived from the same data set that was used in Fig. 5. Note that as the window size increases, the metric value approaches 1. Our experiments indicate that a window around $W = 7$ works well for a variety of images and distortions.

V. USING SSIM TO ASSESS VIDEO QUALITY

A. Motivation for Exploring Space of Realistic Distortions

The objective quality metrics that we examine in this paper are intended to provide a measure of perceptual similarity between a distorted and a reference image (perceptual image fidelity). Of course, since the images are intended to be viewed
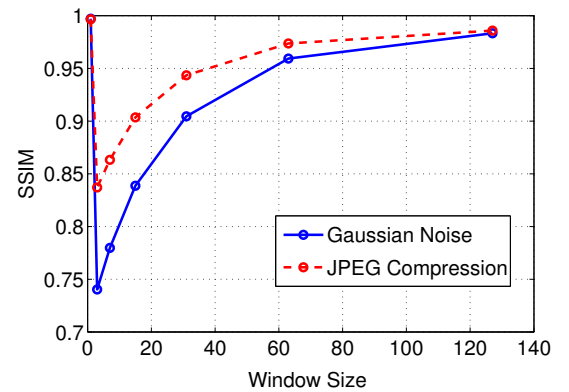

Fig. 4. Mean SSIM versus window size for the same data set in Fig. 5.

by human observers, the metric predictions should agree with subjective evaluations. Subjective evaluation studies collect opinion scores for a database of distorted images and use statistical analysis to compute a mean opinion score (MOS) for each image. The MOS data can then be compared with the quality metric predictions to validate the effectiveness of the metric (e.g., see [19]). The selection of the database of distorted images is critical for the success of such subjective evaluations, which are quite cumbersome and expensive, as
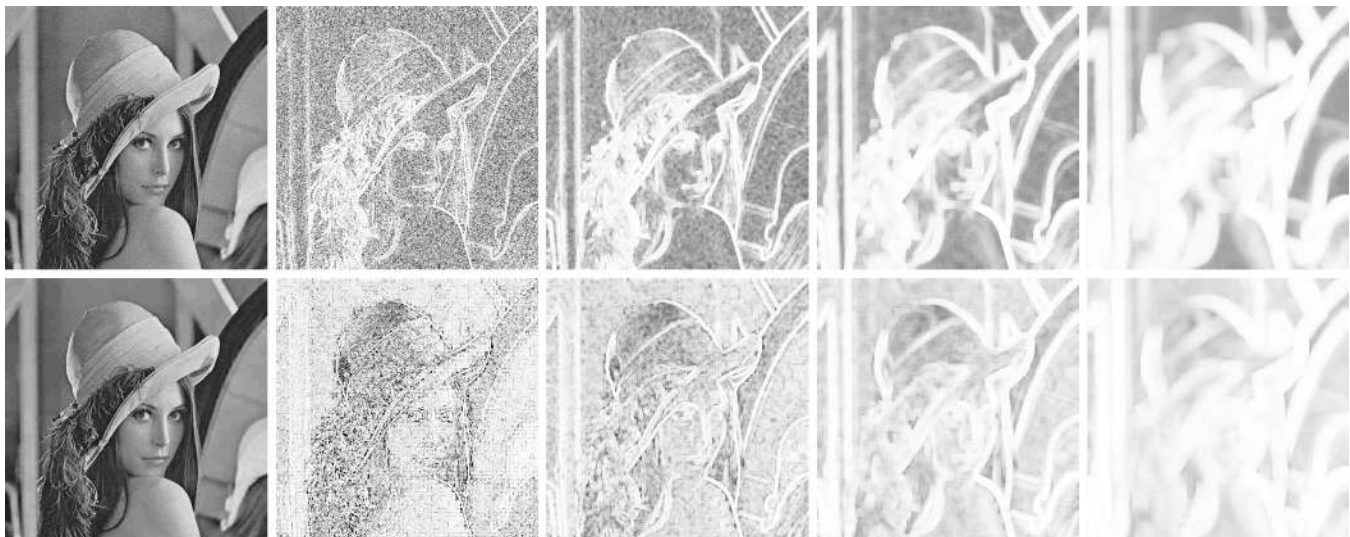
Fig. 5. SSIM error maps for different window sizes. The top row shows the "Lena" image distorted with white Gaussian noise, then error maps for window sizes of 3, 7, 15, and 31. The bottom row shows "Lena" degraded with JPEG compression and corresponding error maps. In the error maps, darker regions represent higher measured distortion.

an inadequate selection can lead to inconclusive results. In addition, when designing a metric, an appropriate choice of distorted test images can significantly improve the process of metric design, by providing intuitive insights into ways for improving the metric and eliminating the need for lengthy subjective evaluations after each modification of the metric parameters.

An alternate way of evaluating image quality metrics was proposed in [20], where the authors suggest a *stimulus synthesis* approach. The idea is to compare two metrics by exploring the image space of metric A while holding metric B constant. This is done via a gradient descent algorithm, producing the "best" and "worst" images in terms of metric A for constant metric B and vice-versa. This allows for efficient evaluation of metrics because the observer only has to look at a few images (the best and worst) to find weaknesses of a metric. A limitation of this approach, however, is that the iterative approach produces distortions that are unlikely to be encountered in compressed video communication systems.

We propose a method of evaluating quality metrics in which we explore the space of realistic distortions that are likely in video compression and communication applications. In our method we hold metric A constant, then examine the results given by metric B with different distortions. This approach provides efficient and valuable intuition for the further improvement of an image quality metric. In the remainder of this section, we present the specifics of the proposed realistic distortion models, define the space of realistic distortions, and discuss specific examples of the performance of different metrics for degraded images in the distortion space.

B. Coding and Concealment Distortion Models

A variety of distortions can be created in video transmission applications due to source coding or packet loss and concealment. Lossy video compression distorts the video before it is transmitted. If the channel is lossy, the error concealment techniques necessary to reconstruct the video introduce further

distortion. In this paper, we propose a set of realistic models for the distortions that are likely in a video transmission system.

We develop models that can be used to simulate video coding and concealment artifacts using a still image. The advantage of this approach is that it allows the study of video distortions in detail without the complexity of evaluating the performance of an entire video compression, transmission, and concealment system. In addition, it provides more flexibility in isolating and controlling the severity of specific types of distortions (e.g., blocking versus blurring), which allows us to develop intuition about the effect each distortion type has on a quality metric.

Error concealment is often necessary in applications such as real-time video streaming over a lossy network, where packets are lost or arrive too late to be useful. The most elementary error concealment approach reconstructs lost image data using spatial interpolation, which may result in significant blurring [21]. Improved quality is possible using spatiotemporal approaches that reconstruct by estimating the lost motion vector and then substituting displaced patches of image data from past video frames. Typically, any part of a compressed video stream can be subject to loss, resulting in distortion. For example, loss of motion vectors may lead to spatial block shift distortions, while loss of DC coefficients could create shifts in the mean values of the video blocks affected.

We assume a block-based compression technique where the basic units susceptible to distortion are square macroblocks (MB) with $N \times N$ pixels (typically $N = 16$). The first distortion arises from compression, which is typically applied on smaller 8×8 blocks. A number of MBs M (sometimes called a *slice*) are then grouped into a packet. For simplicity, we use a straightforward channel model, whereby a packet is lost with probability P_k. We also include a parameter that controls the grouping of distortions within the MBs that make up a single lost packet.

Once it is determined that a macroblock is lost, the model

Fig. 6. Example frame showing distortions created in a spatial replacement error concealment approach.

applies a distortion consistent with the concealment technique used by the receiver. We consider the following types of distortion: block blur (spatial interpolation), block spatial shift (temporal replacement), block intensity shift (loss of DC coefficient), and JPEG compression (DCT coefficient quantization).

Block blur is modeled as the convolution of the MB image patch $\mathbf{x}$ with a 2-D $(N+1) \times (N+1)$ smoothing filter $\mathbf{f}$

$$x_{\mathrm{blur}}(\mathbf{x}, \mathbf{f}) = x(n_1, n_2) * f(n_1, n_2), \tag{14}$$

where "$*$" is the 2-D convolution operator. This is a model of simple spatial interpolation concealment techniques that may use bilinear interpolation to recreate the lost macroblock data. Of course more sophisticated spatial interpolation techniques exist, but this provides a first order approximation. Moreover, the Gaussian approximation makes it possible to control the degree of blurring.

Block spatial shift is modeled as a uniform distribution of spatial shifts with a maximum shift of $\pm B$ pixels. The spatially shifted image patch is described as

$$x_{\mathrm{shift}}(\mathbf{x}, b) = x(n_1 + b_1, n_2 + b_2), \tag{15}$$

where b_1 and b_2 are independent random values chosen from the uniform distribution in the interval $(-B, B)$. This models the effect of temporal replacement concealment techniques that can be used when motion vectors are lost or corrupted, such as motion-compensated temporal replacement [21].

Block intensity shift is modeled as a uniform distribution of block mean shifts with a maximum shift factor of $\pm L$, where $\pm L$ is varied from 0 to 1. The block shifted patch is defined as

$$x_{\mathrm{level}}(\mathbf{x}, L) = x(n_1, n_2) + s, \tag{16}$$

where s is a random value chosen from the uniform distribution in the interval $-256L, 256L$ for 8-bit grayscale images. This models the distortion that can occur when a DC coefficient is reconstructed with some error.

Finally, we model source coder distortion as the DCT coefficient quantization that results when we apply JPEG compression [17]. This generates 8x8 block distortion across the entire image and serves as a model of what might occur in a communication system where source bit rate is sacrificed in order to achieve improved error resilience. In the experiments

below, we use JPEG with a perceptual quantization matrix weighted for a viewing distance of six image heights.

Many communication systems produce distortions that can be modeled with these four distortions. For example, Fig. 6 shows a single frame from a video with spatial translation and intensity shift distortions that occur when using a spatial replacement error concealment approach. This example corresponds to the techniques described in [22].

In all of the above cases, the model parameters (probability of packet loss, parameters of the blur filter, and spatial shift and intensity shift distributions) can be selected to match any given value of any of the objective quality metrics under consideration.

Now that we have modeled realistic distortions, we can perform experiments to find out if SSIM quality measurements agree with the intuition that some distortions are more visible than others. In the following experiments, we assume a MB size of $N = 16$, and as we pointed out in Section III-B, we compute the CWSSIM metric using a two-scale, 16-orientation steerable filter decomposition and average over all the subbands except the high pass.

C. The Space of Realistic Coding Distortions

The distortion model developed in section V-B can be used as a tool for exploring the performance of SSIM metrics over a range of distortions. The resulting image data generated by running the distortion model can be viewed as an exploration of the multidimensional image space of realistic distortions in a video communications system. Fig. 7 shows some example data generated from the "Lena" test image with a window size of $W = 7$. The spatial resolution of the test image is 256×256 pixels, and the MB size for the error concealment distortions is 16×16 pixels. The columns have approximately equal PSNR with decreasing error from left to right. The metric values for the various quality metrics are also given on the right of each image.

A few observations are in order. First, overall, image quality increases from left to right. This means that, when we compare images with the same type of distortion, PSNR is a good predictor of image quality, and so are all the other metrics. Second, one can argue that the block blur artifact is the most objectionable artifact. This is because it destroys the image structure within the blurred region and additionally creates obvious block edges. Another observation is that the relative perceptual quality for some of the distortions changes as we move to higher PSNRs. For example, at 24.6 dB, JPEG compression is one of the most objectionable, but as the PSNR increases, it becomes less objectionable relative to the other distortions. Finally, the perceptual quality for some of the distortions, like white noise and JPEG compression, changes significantly with viewing distance, while for others, like the blur and intensity shift, it is less sensitive to the viewing distance. (The reader is encouraged to try viewing the PDF file at different zoom settings.)

Here we should point out that our goal is not to conduct systematic subjective tests to establish the validity of the metric results, but to demonstrate that the proposed distortion

Fig. 7. Example realistic distortion images where columns have approximately equal PSNR. From top to bottom row, the distortions are (1) block blur, (2) block intensity shift, (3) JPEG compression, (4) block spatial shift, and (5) additive white noise. The metric values to the right of the images correspond to PSNR, Perceptual PSNR, SSIM, CWSSIM, and WCWSSIM.

models can be used to explore the advantages and limitations of the different quality metrics. Thus, we rely on informal subjective evaluations and concentrate on obvious differences rather than subtle details.

Fig. 8 shows the full-size images that correspond to the middle column of Fig. 7. Here, one can argue that the spatial shift has the best overall subjective quality, followed by the noise and intensity shift, and the JPEG and blur artifacts are the most objectionable. Again, we rely on informal subjective evaluations, whereby observers quickly see the distortion in the JPEG image but typically have to look very closely to even find the problems in the spatially distorted image (e.g., block shifts near the mouth, the hair on the right, the ribbon on the hat, etc.). Note that the SSIM and CWSSIM metrics predict the relative performance of the spatial shift image and the JPEG compression, but give the best score to the level shift, and the worst score to the noise, both at odds with the subjective evaluations. The P-PSNR, on the other hand, gives the JPEG and noise images the highest scores, also contrary to the subjective evaluations. Comparing the

SSIM and CWSSIM metric predictions, we see that the latter performs a bit better, giving relatively higher values for the spatial shift. Here we should point out that image content can have a significant effect on the perceived distortions, for example, the distortion is more noticeable when it affects the eyes or some other salient part of the image. Barring this type of situation, however, our extensive experiments indicate that the results we present here are fairly representative of most situations.

Fig. 9 shows another example with a 512×512 image; this is a segment of the JPEG2000 "Bike" test image. The MB size is the same as in Fig. 8, and the other distortion parameters have similar values; the PSNR is equal to 27 dB. Here again the the spatial shift has the best overall quality, but now the intensity shift and blurring artifacts are the most objectionable. Note that, even though the PSNR is lower, the JPEG artifacts and white noise distortion become less objectionable, due to the higher spatial resolution. (For the same viewing distance, the viewing angle in pixels/degree has been reduced by a factor of two.) Note also that the smaller

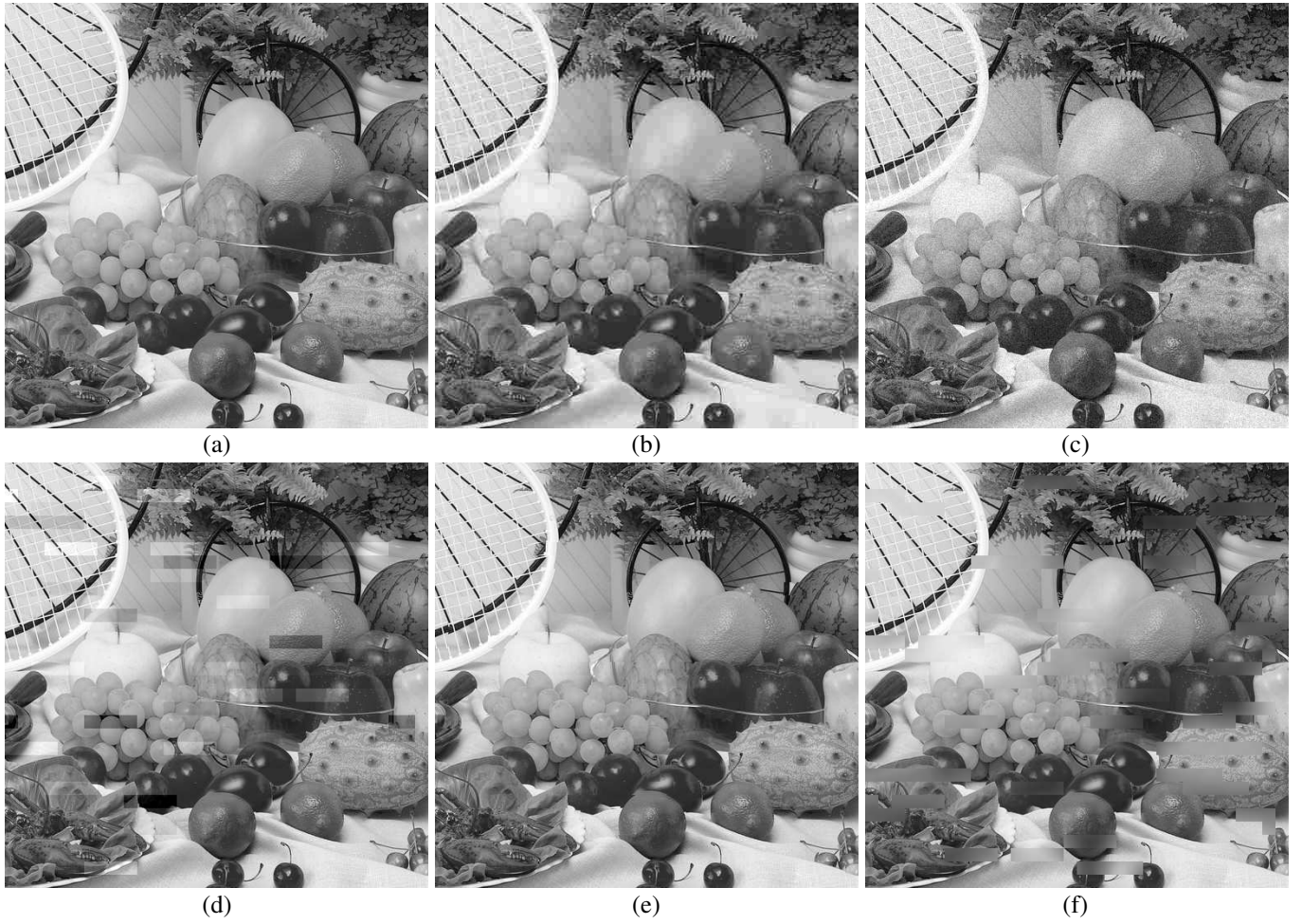| Distortion	P-PSNR	SSIM	CWSSIM	HP	L1	L2	L3	LP	WCWSSIM
(b) JPEG	35.70	0.837	0.757	0.152	0.489	0.777	0.944	0.999	0.867
(c) Noise	35.88	0.711	0.751	0.290	0.493	0.793	0.967	1.000	0.878
(d) Level Shift	32.69	0.971	0.957	0.930	0.941	0.926	0.939	0.991	0.954
(e) Spatial Shift	35.47	0.925	0.946	0.867	0.905	0.933	0.970	0.998	0.963
(f) Blur	33.29	0.923	0.896	0.828	0.851	0.846	0.879	0.995	0.907
Weights				0.000	0.127	0.229	0.306	0.338	

Fig. 8. Degraded images with equal PSNR = 29 dB. (a) Original; (b) JPEG compression; (c) white noise; (d) block intensity shifts; (e) block spatial shifts; (f) block blur. These images correspond to the middle column of Fig. 7. The values for different metrics are given in the table.

relative size of the concealed packets makes the blurring and intensity shift artifacts a bit less objectionable, too. Again, the SSIM and CWSSIM metrics predict the relative performance of the spatial shift image and the JPEG compression, but give the best score to the level shift, and the worst scores to the noise and JPEG compression, both at odds with the subjective evaluations. The P-PSNR, consistent with the earlier observations, gives the JPEG and noise images the highest scores, also contrary to the subjective evaluations. One could argue that the examples in this figure correspond to CCIR601 transmission over packet-lossy networks, while the 256×256 examples of Fig. 8 correspond to the CIF case.

Figs. 8 and 9 demonstrate both advantages and weaknesses of the SSIM and CWSSIM metrics. A significant advantage is that, as we saw above, they predict the relative performance of the spatial shift image and the JPEG compression. These types of distortions are commonly encountered in video compression/transmission applications, where there is a tradeoff between source coding errors and errors due to packet losses

[23]. In a specific example that demonstrates the potential of the SSIM metrics, a video sequence was encoded at a lower rate so that the entire sequence can be transmitted without losses; the resulting PSNR (averaged over the sequence) was 30.0 dB. We then encoded the same sequence at a significantly higher rate, so that a number of blocks had to be dropped and concealed; the resulting average PSNR was 27.5 dB. A representative frame from each implementation is shown in Fig. 10. Despite the packet losses, which result in some incorrect blocks in the decoded image, the visual quality of the second implementation was significantly higher and was correctly predicted by the SSIM and CWSSIM metrics.

A weakness of the SSIM and CWSSIM metrics is the fact that the intensity shift distortions have higher values than the spatial shift distortions. Most human observers make the opposite choice, giving the spatially shifted image higher quality ratings. Another weakness of the SSIM and CWSSIM metrics becomes obvious when we compare the metric predictions for the white noise image in the two figures to those for
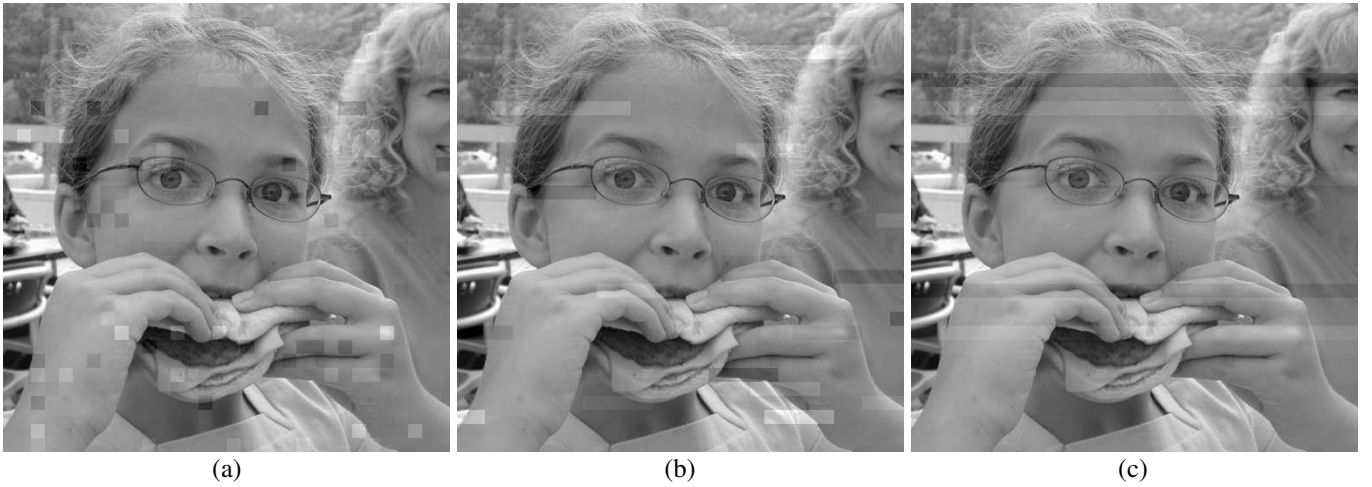
Distortion	P-PSNR	SSIM	CWSSIM	HP	L1	L2	L3	LP	WCWSSIM
(b) JPEG	43.68	0.784	0.830	0.173	0.522	0.834	0.963	1.000	0.890
(c) Noise	43.77	0.709	0.844	0.420	0.588	0.822	0.968	1.000	0.897
(d) Level shift	33.28	0.960	0.949	0.957	0.957	0.940	0.940	0.964	0.950
(e) Spatial Shift	37.21	0.907	0.947	0.855	0.893	0.927	0.972	0.998	0.960
(f) Blur	34.69	0.900	0.903	0.846	0.861	0.870	0.890	0.989	0.920
Weights				0.000	0.127	0.229	0.306	0.338	

Fig. 9.   Degraded images with equal PSNR = 27 dB. (a) Original; (b) JPEG compression; (c) white noise; (d) block intensity shifts; (e) block spatial shifts; (f) block blur. The values for different metrics are given in the table.

Fig. 10.   Video over wireless example: (a) original frame; (b) frame encoded at lower rate without packet losses (PSNR=30.5 dB, CWSSIM=0.88). (c) frame encoded at higher rate with packet losses (PSNR=28.3 dB, CWSSIM=0.90). CWSSIM metric predicts that video on the right is of higher quality in agreement with subjective evaluations.

Distortion	P-PSNR	SSIM	CWSSIM	WCWSSIM
$M = 1$	38.4	0.966	0.929	0.922
$M = 5$	38.2	0.978	0.960	0.954
$M = 32$	38.1	0.978	0.961	0.954

Fig. 11. Effect of MB grouping parameter (M blocks per packet) for equal PSNR (31.9 dB). (a) $M = 1$, (b) 5, and (c) 32 (entire row). The CWSSIM and WCWSSIM values agree with subjective evaluations that image on the left has lower perceptual quality than the other two.

the blurring and intensity shift distortions. The SSIM and CWSSIM values for the white noise distortion are too low, especially for Fig. 9.

The tables associated with Figs. 8 and 9 show a scale-by-scale CWSSIM calculation (over three scales, the low-pass, and the high-pass subbands), each averaged over six angular orientations for the different distortion types.[3] Note that for the white noise and JPEG compression, the value of the metric varies widely from level to level, while for the other distortions the value of the metric does not change as much. Thus, the CWSSIM metric values are expected to change significantly if a different set of subbands is used to compute the metric value. As we will see below, simple averaging is not necessarily the best thing to do.

Another interesting application of the proposed distortion models is depicted in Fig. 11, where we evaluate the effect of MB grouping into M blocks per packet. For approximately equal PSNR (31.9 dB), both SSIM and CWSSIM predict that the images with single MB packets ($M = 1$) have lower quality than the images with packets that contain $M = 5$ MBs or an entire row of MBs ($M = 32$). This result agrees with informal subjective, and indicates that using larger packet sizes to obtain better compression efficiency does not result in any losses in perceptual quality.

D. Perceptually-Weighted CWSSIM

In order to address some of the weaknesses described above, we propose a weighted complex wavelet SSIM implementation (WCWSSIM) that combines the results from multiple wavelet scales. This is similar to the multiscale

---

[3]Note that the CWSSIM metric values are still computed using a two-scale, 16-orientation decomposition, averaged over the two scales and 16 orientations, as we saw in Section III-B.

SSIM approach proposed in [24], but uses the steerable filter decomposition instead of the Gaussian pyramid. In order to avoid cumbersome subjective experiments for determining the subband weights, we loosely base the weight selection on the wavelet threshold measurements in [25]. (See also [1].) For the weight calculation, we assume a viewing distance of six image heights at the 512×512 resolution, which corresponds to 12 image heights at the 256×256 resolution (so that the viewing angle in pixels/degree remains the same). Note that even though the two wavelet decompositions are quite different, they are both octave spaced, and there is an obvious correspondence between subbands: the L3, L2, and L1 subbands of the steerable pyramid correspond to the level 3, 2, and 1 LH (or HL) bands of the wavelet decomposition in [25], respectively, while the LP steerable subband corresponds to the level 3 LL band in [25]. The resulting weights are listed in the tables in Figs. 8 and 9. Note that this is only a rough approximation. More accurate estimates should be based on subjective experiments that determine the base sensitivity of each subband for a given viewing distance [1]. The main point we are trying to make is that the subband weights should depend on frequency content and viewing distance. We should also point out that all we propose is to use weights based on (JND) threshold measurements. Optimizing the weights for suprathresold distortions would be quite difficult, and is beyond the scope of this paper.

Using these weights to combine the results from multiple subbands, we find that this WCWSSIM version of the metric improves the results of Section V-C. Note that in both Figs. 8 and 9, WCWSSIM gives higher quality ratings to the spatial shift than the level shift images, better agreeing with the higher perceived quality of those images. However, while this definitely represents an improvement, it still does not reflect the (considerably) superior quality of the spatial shift images.

We believe that the remaining problems are due to the fact that the complex SSIM insensitivity to translations is limited to small translations, while the HVS can tolerate a lot more. In addition, the SSIM metrics are too insensitive to intensity shifts. The WCWSSIM metric also increases the values for the JPEG and white noise distortions. The metric value for the white noise distortion, however, is still lower than that of the blur, which is definitely of lower perceptual quality.

A block diagram of the perceptually-weighted CWSSIM implementation is shown in Fig. 12. The error calculation consists of the SSIM we reviewed in Section III-A. Note that there is no explicit masking (shown in dotted line); as we saw in Section IV-B the contrast masking is implicit, primarily through the contrast comparison term of (6). Note also, that if we change the error calculation and make it an MSE computation and add the masking term, we get the perceptual metric of (3). Thus, the perceptual weighting of the WCWSSIM metric provides a natural way to unify the structural similarity approach with the traditional JND-based perceptual approaches.
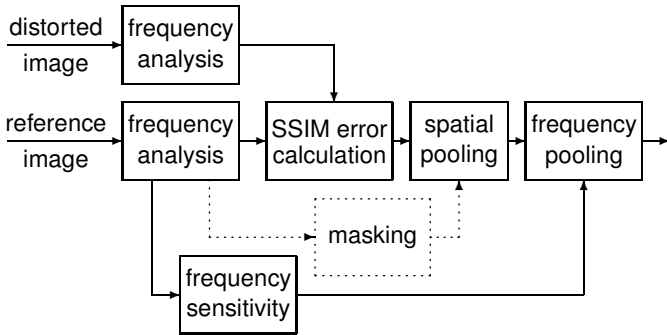

Fig. 12. Perceptual coder

VI. Conclusion

We have examined the use of structural similarity metrics in suprathreshold video compression/transmission applications. In order to better explore the space of distortions, we proposed models for typical distortions encountered in these applications. Our experimental results indicate that, while structural similarity metrics provide significant advantages over traditional approaches, they also have important limitations. We believe that the proposed distortion models will be the key to addressing such limitations. We also found that the translation-insensitive complex wavelet SSIM (CWSSIM) is superior to other SSIM implementations, and proposed a perceptually weighted multi-scale variant of CWSSIM that accounts for viewing distance and provides a natural way to unify the structural similarity approach with the traditional JND-based perceptual approaches.

Acknowledgment

References

[1] T. N. Pappas, R. J. Safranek, and J. Chen, "Perceptual criteria for image quality evaluation," in *Handbook of Image and Video Processing*, 2nd ed., A. C. Bovik, Ed. Academic Press, 2005, pp. 939–959.

[2] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[3] M. P. Eckert and A. P. Bradley, "Perceptual quality metrics applied to still image compression," *Signal Processing*, vol. 70, pp. 177–200, 1998.

[4] T. N. Pappas, T. A. Michel, and R. O. Hinds, "Supra-threshold perceptual image coding," in *Proc. Int. Conf. Image Processing (ICIP-96)*, vol. I, Lausanne, Switzerland, Sept. 1996, pp. 237–240.

[5] J. Chen and T. N. Pappas, "Perceptual coders and perceptual metrics," in *Human Vision and Electronic Imaging VI*, B. E. Rogowitz and T. N. Pappas, Eds., vol. Proc. SPIE Vol. 4299, San Jose, CA, Jan. 2001, pp. 150–162.

[6] R. J. Safranek and J. D. Johnston, "A perceptually tuned sub-band image coder with image dependent quantization and post-quantization data compression," in *Proc. ICASSP-89*, vol. 3, Glasgow, Scotland, May 1989, pp. 1945–1948.

[7] A. B. Watson, "DCT quantization matrices visually optimized for individual images," in *Human Vision, Visual Proc., and Digital Display IV*, J. P. Allebach and B. E. Rogowitz, Eds., vol. Proc. SPIE, Vol. 1913, San Jose, CA, Feb. 1993, pp. 202–216.

[8] S. S. Hemami and M. G. Ramos, "Wavelet coefficient quantization to produce equivalent visual distortion in complex stimuli," in *Human Vision and Electronic Imaging V*, B. E. Rogowitz and T. N. Pappas, Eds., vol. Proc. SPIE Vol. 3959, San Jose, CA, Jan. 2000, pp. 200–210.

[9] M. G. Ramos and S. S. Hemami, "Suprathreshold wavelet coefficient quantization in complex stimuli: psychophysical evaluation and analysis," *J. Opt. Soc. Am. A*, vol. 18, no. 10, pp. 2385–2397, Oct. 2001.

[10] D. M. Chandler and S. S. Hemami, "Additivity models for suprathreshold distortion in quantized wavelet-coded images," in *Human Vision and Electronic Imaging VII*, B. E. Rogowitz and T. N. Pappas, Eds., vol. Proc. SPIE Vol. 4662, San Jose, CA, Jan. 2002, pp. 105–118.

[11] ——, "Effects of natural images on the detectability of simple and compound wavelet subband quantization distortions," *J. Opt. Soc. Am. A*, vol. 20, no. 7, pp. 1164–1180, July 2003.

[12] C. J. van den Branden Lambrecht and O. Verscheure, "Perceptual quality measure using a spatio-temporal model of the human visual system," in *Digital Video Compression: Algorithms and Technologies*, V. Bhaskaran, F. Sijstermans, and S. Panchanathan, Eds., vol. Proc. SPIE, Vol. 2668, San Jose, CA, Jan./Feb. 1996, pp. 450–461.

[13] Z. Wang, A. C. Bovik, and L. Lu, "Why is image quality assessment so difficult?" in *IEEE Int. Conf. on Acoustics, Speech, and Signal Proc.*, vol. IV, 2002, pp. 3313–3316.

[14] Z. Wang, A. C. Bovik, and E. P. Simoncelli, "Structural approaches to image quality assessment," in *Handbook of Image and Video Processing*, 2nd ed., A. C. Bovik, Ed. Academic Press, 2005, pp. 961–974.

[15] Z. Wang and E. P. Simoncelli, "Translation insensitive image similarity in complex wavelet domain," in *IEEE Int. Conference on Acoustics, Speech, and Signal Processing*, vol. II, Philadelphia, PA, 2005, pp. 573–576.

[16] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, "Shiftable multi-scale transforms," *IEEE Trans. Inform. Theory*, vol. 38, no. 2, pp. 587–607, Mar. 1992.

[17] J. L. M. William B. Pennebaker, Ed., *JPEG: Still Image Data Compression Standard*. New York: Van Nostrand Reinhold, 1993.

[18] Z. Wang, L. Lu, and A. Bovik, "Video quality assessment based on structural distortion measurement," pp. 121–132, Feb. 2004. [Online]. Available: http://citeseer.ist.psu.edu/wang04video.html

[19] Video Quality Experts Group (VQEG), http://www.its.bldrdoc.gov/vqeg/.

[20] Z. Wang and E. P. Simoncelli, "Stimulus synthesis for efficient evaluation and refinement of perceptual image quality metrics," in *Human Vision and Electronic Imaging IX*, B. E. Rogowitz and T. N. Pappas, Eds., vol. Proc. SPIE, Vol. 5292, San Jose, CA, Jan. 2004, pp. 99–108. [Online]. Available: citeseer.ist.psu.edu/article/wang04stimulus.html

[21] Y. Wang, J. Ostermann, and Y. Zhang, *Video Processing and Communicaitons*. Upper Saddle River, New Jersey: Prentice Hill, 2002.

[22] Y. Eisenberg, T. N. Pappas, R. Berry, and A. K. Katsaggelos, "Minimizing transmission energy in wireless video communications," in *Proc. Int. Conf. Image Processing (ICIP-01)*, vol. 1, Thessaloniki, Greece, Oct. 2001, pp. 958–961.

[23] P. Pahalawatta, R. Berry, T. Pappas, and A. Katsaggelos, "Content-aware resource allocation and packet scheduling for video transmission over wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 4, pp. 749–759, May 2007.

[24] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," in *37th IEEE Asilomar Conf. on Signals, Systems and Computers*, vol. 2, Pacific Grove, CA, Nov. 2003, pp. 1398–1402.

[25] A. B. Watson, G. Y. Yang, J. A. Solomon, and J. Villasenor, "Visual thresholds for wavelet quantization error," in *Proc. SPIE, vol. 2657, Human Vision and Electronic Imaging*, San Jose, CA, Jan. 29–Feb. 1, 1996, pp. 382–392.

PLACE PHOTO HERE

Alan C. Brooks Alan Brooks earned the B.S. degree at Bradley University in 2001 and the M.S. degree in electrical engineering and computer science at Northwestern University in 2006. His research interests include image and video coding and quality metrics that account for human perception. Alan has worked at Northrop Grumman corporation for over seven years, where he has designed systems in the image/signal processing, EO/IR, user interface, and inertial navigation disciplines. He has also led test and development teams testing out new technologies at many of the Air Force bases throughout the United States.

PLACE PHOTO HERE

Xiaonan Zhao Xiaonan Zhao received the B.S. degree in electronic and information engineering from Zhejiang University, Hangzhou in 2005, and the M.S. degree in electrical engineering and computer science from Northwestern University, Evanston, IL in 2007. She is currently a Ph.D. student in electrical engineering and computer science at Northwestern University.

PLACE PHOTO HERE

Thrasyvoulos N. Pappas (M'87, SM'95, F'06) received the S.B., S.M., and Ph.D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology, Cambridge, MA, in 1979, 1982, and 1987, respectively. From 1987 until 1999, he was a Member of the Technical Staff at Bell Laboratories, Murray Hill, NJ. In 1999, he joined the Department of Electrical and Computer Engineering (now EECS) at Northwestern University as an associate professor. His research interests are in image and video quality and compression, perceptual models for image processing, image and video analysis, model-based halftoning, and multimedia signal processing.

Dr. Pappas has served as an elected member of the Board of Governors of the Signal Processing Society of IEEE (2004-2007), chair of the IEEE Image and Multidimensional Signal Processing Technical Committee, associate editor of the IEEE Transactions on Image Processing, and technical program co-chair of ICIP-01 and the Symposium on Information Processing in Sensor Networks (IPSN-04). Dr. Pappas is a Fellow of SPIE. Since 1997 he has been co-chair of the SPIE/IS&T Conference on Human Vision and Electronic Imaging. He has also served as co-chair of the 2005 SPIE/IS&T Electronic Imaging Symposium.