

# UNIFYING ANALYSIS OF FULL REFERENCE IMAGE QUALITY ASSESSMENT

Kalpna Seshadrinathan and Alan C. Bovik

Dept. of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX, U.S.A.

## ABSTRACT

This paper studies two increasingly popular paradigms for image quality assessment - Structural SIMilarity (SSIM) metrics and Information Fidelity metrics. The relation of the SSIM metric to Mean Squared Error and Human Visual System (HVS) based models of quality assessment are studied. The SSIM model is shown to be equivalent to models of contrast gain control of the HVS. We study the information theoretic metrics and show that the Information Fidelity Criterion (IFC) is a monotonic function of the structure term of the SSIM index applied in the sub-band filtered domain. Our analysis of the Visual Information Fidelity (VIF) criterion shows that improvements in VIF include incorporation of a contrast comparison, in addition to the structure comparison in IFC. Our analysis attempts to unify quality metrics derived from different first principles and characterize the relative performance of different QA systems.

**Index Terms**— Image quality, Structural Similarity, Visual Information Fidelity, Quality assessment.

## 1. INTRODUCTION

There is increasing interest in objective methods of quality computation of images and since the ultimate receiver in a large number of applications is a human observer, a large body of work in the literature has focused on computing the quality of an image *as seen by a human observer*. Full reference quality assessment (QA) algorithms assume the availability of a “perfect” quality reference signal and a large body of work on quality assessment has contributed to the study of this problem.

Different QA algorithms have been proposed in the literature and the performance of these algorithms is usually evaluated by constructing a database of distorted images and obtaining ground truth quality scores for these images from human observers. Statistical metrics such as the correlation coefficient are then used as indicators of performance after fitting the algorithm predictions to the subjective scores using a monotonic function. Such analysis has certain drawbacks since the results depend on the form of the fitting function, optimization methods used in fitting and the choice of free parameters in the QA system. In this paper, we analyze two recently developed, popular image QA paradigms known as the Structural SIMilarity paradigm (SSIM) [1, 2] and the Visual Information Fidelity (VIF) paradigm [3, 4] in a general probabilistic framework. We believe that the analytic characterization of these systems in this paper avoids the drawbacks of experimental validation and deepens our understanding of these metrics. We will also attempt to relate the SSIM and VIF QA paradigms to more traditional QA metrics: the Mean Squared Error (MSE) and HVS based image quality metrics.

We begin by introducing the notation that we will use throughout this paper. Let  $F(\mathbf{i})$  denote a random variable that models a pixel at spatial location  $\mathbf{i}$  in the reference image. Similarly, let  $G(\mathbf{i})$  denote

a random variable that models the corresponding pixel from the test image. Let  $\tilde{f}(\mathbf{i})$  and  $\tilde{g}(\mathbf{i})$  denote the reference and test images respectively. We define a sequence of  $N$  vectors  $\mathbf{f}(\mathbf{i})$  and  $\mathbf{g}(\mathbf{i})$ , where  $\mathbf{f}(\mathbf{i})$  is composed of  $N$  elements of  $\tilde{f}(\mathbf{i})$  spanned by the window  $B_1$  and similarly for  $\mathbf{g}(\mathbf{i})$ . To index each element of  $\mathbf{f}(\mathbf{i})$ , we use the notation  $\mathbf{f}(\mathbf{i}) = [f_1(\mathbf{i}), f_2(\mathbf{i}), \dots, f_N(\mathbf{i})]^T$ . Although the window  $B_1$  can be of any shape, in practice, it usually spans a rectangular region of connected pixels. Consider the linear shift-invariant filtering of  $f(\mathbf{i})$  and  $g(\mathbf{i})$  by a family of two-dimensional sub-band kernels, denoted  $h(\mathbf{i}, k)$ , where  $k$  indexes over each filter in the family. Let  $X(\mathbf{i}, k)$  and  $Y(\mathbf{i}, k)$  denote random variables that model the coefficient at spatial location  $\mathbf{i}$  obtained by filtering the reference and test image patches with the  $k^{\text{th}}$  filter  $h(\mathbf{i}, k)$  respectively. Let  $\tilde{x}(\mathbf{i}, k)$  and  $\tilde{y}(\mathbf{i}, k)$  denote the coefficients of the  $k^{\text{th}}$  sub-band of the reference and test images respectively. Finally, define  $M$  dimensional vectors  $\mathbf{x}(\mathbf{i}, k)$  and  $\mathbf{y}(\mathbf{i}, k)$  that contain  $M$  coefficients of  $\tilde{x}(\mathbf{i}, k)$  and  $\tilde{y}(\mathbf{i}, k)$  spanned by a window  $B_2$  respectively. We use similar notation to index each element of  $\mathbf{x}(\mathbf{i}, k)$  and  $\mathbf{y}(\mathbf{i}, k)$ , i.e.  $\mathbf{x}(\mathbf{i}, k) = [x_1(\mathbf{i}, k), x_2(\mathbf{i}, k), \dots, x_M(\mathbf{i}, k)]^T$  etc.

## 2. MSE AND HVS BASED METRICS

The MSE between image patches  $\mathbf{f}(\mathbf{i})$  and  $\mathbf{g}(\mathbf{i})$  is defined by:

$$\text{MSE}[\mathbf{f}(\mathbf{i}), \mathbf{g}(\mathbf{i})] = \frac{1}{N} \sum_{j=1}^N [f_j(\mathbf{i}) - g_j(\mathbf{i})]^2$$

HVS based metrics use psychophysical measurements of the characteristics of the vision system to compute visual quality. QA models based on the HVS are rather elaborate and model several different aspects of the HVS, but we concentrate on models of contrast masking in the HVS. Contrast masking refers to the reduction in visibility of a signal component (distortion) due to the presence of another signal component of similar frequency and orientation (image content) in a local spatial neighborhood. Contrast masking has been modeled using contrast gain control which is a mechanism that allows a neuron in the HVS to adjust its response to the ambient contrast, thereby keeping the neural responses within their permissible dynamic range. HVS based metrics typically decompose the image using a linear sub-band decomposition and gain control is modeled in the sub-band decomposed domain. Contrast gain control models usually take the form of a divisive normalization, where the response of a neuron has an accelerating nonlinearity, but is also inhibited divisively by the response of a local pool of neurons [5, 6, 7]. As an example, Teo and Heeger [6] use the following model to define the response of a neuron to a signal  $\tilde{x}(\mathbf{i}, k)$ .

$$R[\tilde{x}(\mathbf{i}, k)] = \kappa \frac{\tilde{x}(\mathbf{i}, k)^2}{\sum_{k \in K_1} \tilde{x}(\mathbf{i}, k)^2 + C} \quad (1)$$

Summation over the sub-bands in the denominator is only carried out over those sub-bands with the same frequency, but different ori-

Thanks to the National Science Foundation for funding.

entations.  $\kappa$  restricts the dynamic range of the response and  $C$  is a saturation constant that stabilizes the response when the denominator is small. The saturation constant determines the normalized response at low signal energies and is important because the masker (image content) has little or no effect upon the detection of the signal (distortions) in regions of low image/masker energy. The saturation constant adjusts the contrast masking normalization for this effect in this regime. This observation will prove useful in our discussion of the structural similarity index next.

### 3. STRUCTURAL SIMILARITY METRICS

The structural similarity paradigm hypothesizes that the visual quality of a given image is related to the loss of *structural information* in an image [2, 1]. Structure, by definition, is independent of the illumination of an image and the structure is computed after normalizing the image patches for mean luminance and contrast. The SSIM index between image patches  $\mathbf{f}(\mathbf{i})$  and  $\mathbf{g}(\mathbf{i})$  is then defined as a function of a luminance comparison term  $l[\mathbf{f}(\mathbf{i}), \mathbf{g}(\mathbf{i})]$ , contrast comparison term  $c[\mathbf{f}(\mathbf{i}), \mathbf{g}(\mathbf{i})]$  and a structure comparison term  $s[\mathbf{f}(\mathbf{i}), \mathbf{g}(\mathbf{i})]$ .

$$\text{SSIM}[\mathbf{f}(\mathbf{i}), \mathbf{g}(\mathbf{i})] = l[\mathbf{f}(\mathbf{i}), \mathbf{g}(\mathbf{i})]c[\mathbf{f}(\mathbf{i}), \mathbf{g}(\mathbf{i})]s[\mathbf{f}(\mathbf{i}), \mathbf{g}(\mathbf{i})] \quad (2)$$

$$l[\mathbf{f}(\mathbf{i}), \mathbf{g}(\mathbf{i})] = \frac{2\mu_{\mathbf{f}(\mathbf{i})}\mu_{\mathbf{g}(\mathbf{i})} + C_1}{\mu_{\mathbf{f}(\mathbf{i})}^2 + \mu_{\mathbf{g}(\mathbf{i})}^2 + C_1} \quad (3)$$

$$c[\mathbf{f}(\mathbf{i}), \mathbf{g}(\mathbf{i})] = \frac{2\sigma_{\mathbf{f}(\mathbf{i})}\sigma_{\mathbf{g}(\mathbf{i})} + C_2}{\sigma_{\mathbf{f}(\mathbf{i})}^2 + \sigma_{\mathbf{g}(\mathbf{i})}^2 + C_2} \quad (4)$$

$$s[\mathbf{f}(\mathbf{i}), \mathbf{g}(\mathbf{i})] = \frac{\sigma_{\mathbf{f}(\mathbf{i})\mathbf{g}(\mathbf{i})} + C_3}{\sigma_{\mathbf{f}(\mathbf{i})}\sigma_{\mathbf{g}(\mathbf{i})} + C_3} \quad (5)$$

Here,  $\mu_{\mathbf{f}(\mathbf{i})}$  and  $\mu_{\mathbf{g}(\mathbf{i})}$  are the sample means of  $\mathbf{f}(\mathbf{i})$  and  $\mathbf{g}(\mathbf{i})$  and  $\sigma_{\mathbf{f}(\mathbf{i})}^2$  and  $\sigma_{\mathbf{g}(\mathbf{i})}^2$  are their sample variances.  $\sigma_{\mathbf{f}(\mathbf{i})\mathbf{g}(\mathbf{i})}$  is the sample covariance between  $\mathbf{f}(\mathbf{i})$  and  $\mathbf{g}(\mathbf{i})$ .

$$\mu_{\mathbf{f}(\mathbf{i})} = \frac{1}{N} \sum_{j=1}^N f_j(\mathbf{i}), \sigma_{\mathbf{f}(\mathbf{i})}^2 = \frac{1}{N} \sum_{j=1}^N (f_j(\mathbf{i}) - \mu_{\mathbf{f}(\mathbf{i})})^2 \quad (6)$$

$$\sigma_{\mathbf{f}(\mathbf{i})\mathbf{g}(\mathbf{i})} = \frac{1}{N} \sum_{j=1}^N (f_j(\mathbf{i}) - \mu_{\mathbf{f}(\mathbf{i})})(g_j(\mathbf{i}) - \mu_{\mathbf{g}(\mathbf{i})}) \quad (7)$$

The initial metric based on structural similarity, known as the Universal Quality Index (UQI), is also defined by (2-5), with  $C_1 = C_2 = C_3 = 0$ . The constants  $C_1, C_2, C_3$  were included to obtain the improved SSIM index to avoid instability when the denominators of the luminance, contrast and structure comparison terms were too small. Although the SSIM index is defined by three terms, the structure term in the SSIM index is most important since variations in luminance and contrast of an image does not affect visual quality as much as structural distortions [1].

Observe that the estimate of the correlation coefficient in (5) and (7) implicitly assumes that  $[\mathbf{f}(\mathbf{i}), \mathbf{g}(\mathbf{i})]$  are independent samples from a Gaussian distribution, since (5) corresponds to a Maximum Likelihood (ML) estimate of the correlation coefficient under these assumptions. This implicit assumption in defining (5) means that the structure term of the SSIM index does not incorporate any information regarding the distribution of the underlying pixels. To overcome this drawback, we define a *probabilistic* SSIM index between random variables  $F(\mathbf{i})$  and  $G(\mathbf{i})$ , whose structure term is given by:

$$\rho[F(\mathbf{i}), G(\mathbf{i})] = \frac{\text{Cov}[F(\mathbf{i}), G(\mathbf{i})]}{\sqrt{\text{Var}[F(\mathbf{i})]}\sqrt{\text{Var}[G(\mathbf{i})]}}$$

where  $\rho[F(\mathbf{i}), G(\mathbf{i})]$  is the correlation coefficient between the random variables. The SSIM index is then just a special case of this more general probabilistic SSIM index obtained when the  $[F(\mathbf{i}), G(\mathbf{i})]$  are jointly Gaussian distributed.

Note that the use of the correlation coefficient in the SSIM index implies measurement of the *linear dependence* between them as a measure of visual quality. In Section 4, this observation will aid us in understanding the relation between the SSIM and VIF metrics.

#### 3.1. Relation to MSE

We define normalized random variables:

$$F'(\mathbf{i}) = \frac{F(\mathbf{i}) - \mathbf{E}[F(\mathbf{i})]}{\sqrt{\text{Var}[F(\mathbf{i})]}}, G'(\mathbf{i}) = \frac{G(\mathbf{i}) - \mathbf{E}[G(\mathbf{i})]}{\sqrt{\text{Var}[G(\mathbf{i})]}} \quad (8)$$

where  $\mathbf{E}$  stands for the expectation operator. We now observe that:

$$\mathbf{E}[(F'(\mathbf{i}) - G'(\mathbf{i}))^2] = 2(1 - \rho[F(\mathbf{i}), G(\mathbf{i})]) \quad (9)$$

It is fairly straightforward to show that the relation described by (9) holds for the estimates of the correlation coefficient and MSE as well:

$$\begin{aligned} \text{MSE} \left( \frac{\mathbf{f}(\mathbf{i}) - \mu_{\mathbf{f}(\mathbf{i})}}{\sigma_{\mathbf{f}(\mathbf{i})}}, \frac{\mathbf{g}(\mathbf{i}) - \mu_{\mathbf{g}(\mathbf{i})}}{\sigma_{\mathbf{g}(\mathbf{i})}} \right) &= 2 \left( 1 - \frac{\sigma_{\mathbf{f}(\mathbf{i})\mathbf{g}(\mathbf{i})}}{\sigma_{\mathbf{f}(\mathbf{i})}\sigma_{\mathbf{g}(\mathbf{i})}} \right) \\ &= 2\{1 - s[\mathbf{f}(\mathbf{i}), \mathbf{g}(\mathbf{i})]\} \end{aligned} \quad (10)$$

where  $C_3 = 0$  in (10).

Thus, the structure term in the SSIM index computes an MSE between normalized image patches. A different relation between SSIM and the MSE between the original variables ( $\mathbf{f}(\mathbf{i})$  and  $\mathbf{g}(\mathbf{i})$ ) has been reported [8]. Our analysis describes the relation between SSIM and the MSE between *normalized* variables, a distinction that is significant since we attempt to cast the role of this normalization as accounting for contrast masking effects in the HVS as described next.

#### 3.2. Relation to HVS based metrics

The normalized variables described by (8) immediately remind us of contrast gain control models used in HVS based metrics, described in Section 2. In (9), the image pixels are normalized by their standard deviation in a local neighborhood before computing a Minkowski error metric. Thus, the contrast gain control model used in SSIM can be defined by:

$$R[f_j(\mathbf{i})] = \frac{f_j(\mathbf{i}) - \mu_{\mathbf{f}(\mathbf{i})}}{\sqrt{\frac{1}{N} \sum_{j=1}^N (f_j(\mathbf{i}) - \mu_{\mathbf{f}(\mathbf{i})})^2}} \quad (11)$$

We wish to emphasize the fact that the SSIM indices perform this normalization in the *image pixel* domain. However, contrast masking in the HVS is a phenomenon that occurs in a frequency-orientation decomposed domain. For example, the masking effect is maximum when the spatial frequency of the masker and the signal are identical. However, the SSIM metrics will not be able to account for such effects. The analysis here suggests that applying the SSIM index in the sub-band filtered domain would result in better performance. Improved versions of the SSIM index that use such frequency decomposition have been proposed [9, 10] and our analysis of the SSIM index within a contrast gain control framework helps us understand the reasons for the improved performance of these metrics. We will see in Section 4 that the information theoretic metrics are also very closely related to multi-scale structural similarity models.

Relating the SSIM metrics to contrast masking models in HVS based design also provides insights on the need for the constant  $C_3$  in (5).  $C_3$  in SSIM plays a similar role as the saturation constants in HVS based metrics since contrast masking effects are minimal in low signal energy regions. This interpretation helps explain the reasons for the improved performance of the SSIM index over UQI. Interestingly, the square of the response of the SSIM contrast gain control model defined by (11) is equal to response of the Teo and Heeger gain control model defined by (1), if the same inhibitory pool of neurons is used and with the inclusion of a saturation constant in the SSIM model. Finally, it has been observed previously that the standard deviations in the denominator of the SSIM index reflect masking and that the constant  $C_3$  attempts to account for the visibility of distortions when these standard deviations are small [11]. Our analysis supports these observations and explicitly links SSIM and contrast masking models.

## 4. INFORMATION FIDELITY METRICS

### 4.1. The IFC Metric

In the information theoretic approach to quality assessment, the test image is assumed to be the result of the reference image passing through a distortion channel and the visual quality is hypothesized to be related to the *capacity* of this communication channel [3, 4]. The sub-band filtered coefficients of the reference image are modeled as random variables using natural scene statistical models. The preliminary version of the information theoretic framework, known as the Information Fidelity Criterion (IFC) [3], uses the scalar Gaussian Scale Mixture (GSM) model [12, 13] and each scalar coefficient is modeled as a random variable:

$$X(\mathbf{i}, k) = Z(\mathbf{i}, k)U(\mathbf{i}, k)$$

where  $Z(\mathbf{i}, k)$  is a gain field and  $U(\mathbf{i}, k)$  is assumed to be an Additive White Gaussian Noise (AWGN) field of unit variance. The distortion channel is modeled using:

$$Y(\mathbf{i}, k) = \beta(\mathbf{i}, k)X(\mathbf{i}, k) + V(\mathbf{i}, k)$$

where  $\beta(\mathbf{i}, k)$  is the deterministic channel gain and  $V(\mathbf{i}, k)$  is AWGN of variance  $\sigma_v(k)^2$ . The IFC index between  $X(\mathbf{i}, k)$  and  $Y(\mathbf{i}, k)$  is then given by the mutual information between these random variables conditioned on  $Z(\mathbf{i}, k)$ :

$$\begin{aligned} \text{IFC}[X(\mathbf{i}, k), Y(\mathbf{i}, k)] &= \mathbb{I}[X(\mathbf{i}, k), Y(\mathbf{i}, k) | Z(\mathbf{i}, k)] \\ &= \frac{1}{2} \log_2 \left( \frac{\beta(\mathbf{i}, k)^2 Z(\mathbf{i}, k)^2 + \sigma_v(k)^2}{\sigma_v(k)^2} \right) \end{aligned} \quad (12)$$

We observe that due to the use of a linear channel model, the IFC attempts to capture the linear dependence between  $X(\mathbf{i}, k)$  and  $Y(\mathbf{i}, k)$ . This observation hints at the possibility that IFC may be closely related to SSIM. A well known result in information theory states that the mutual information between two variables is a monotonic function of the Pearson correlation coefficient between them if the variables are jointly Gaussian [14]. This result is applicable in our analysis since the GSM models the reference coefficients as Gaussian when conditioned on  $Z(\mathbf{i}, k)$  and the linear channel model causes the reference and test coefficients to be jointly Gaussian. Here, we prove that the same relation holds between the estimates of these quantities used in IFC and SSIM.

To obtain estimates of  $Z(\mathbf{i}, k)$  and  $\beta(\mathbf{i}, k)$ , we need to consider a local neighborhood of coefficients surrounding the spatial location  $\mathbf{i}$ . In the IFC framework, each sub-band is treated

independently and hence, a local neighborhood is extracted by considering coefficients in the same sub-band at adjacent spatial locations [3]. Consistent with our earlier notation, we denote this local neighborhood extracted using a window  $B_2$  as  $\mathbf{x}(\mathbf{i}, k) = [x_1(\mathbf{i}, k), x_2(\mathbf{i}, k), \dots, x_M(\mathbf{i}, k)]^T$  for the reference image coefficients and  $\mathbf{y}(\mathbf{i}, k) = [y_1(\mathbf{i}, k), y_2(\mathbf{i}, k), \dots, y_M(\mathbf{i}, k)]^T$  for the test image. Let  $\hat{Z}(\mathbf{i}, k)$  denotes an estimate of  $Z(\mathbf{i}, k)$  and similarly for  $\hat{\beta}(\mathbf{i}, k)$ . Since the reference image coefficients are assumed to be zero-mean in the GSM model, the Maximum Likelihood estimate of  $\hat{Z}(\mathbf{i}, k)$  is given by:

$$\hat{Z}(\mathbf{i}, k)^2 = \frac{1}{M} \sum_{j=1}^M x_j(\mathbf{i}, k)^2 = \sigma_{\mathbf{x}(\mathbf{i}, k)}^2 \quad (13)$$

If we assume that the parameters of the channel model are also estimated using the same window  $B_2$ , the least squares estimate of  $\hat{\beta}(\mathbf{i}, k)$  and the noise variance obtained using linear regression are [3]

$$\hat{\beta}(\mathbf{i}, k) = \frac{\sigma_{\mathbf{x}(\mathbf{i}, k)\mathbf{y}(\mathbf{i}, k)}}{\sigma_{\mathbf{x}(\mathbf{i}, k)}^2} \quad (14)$$

$$\hat{\sigma}_v(k)^2 = \sigma_{\mathbf{y}(\mathbf{i}, k)}^2 - \hat{\beta}(\mathbf{i}, k)\sigma_{\mathbf{x}(\mathbf{i}, k)\mathbf{y}(\mathbf{i}, k)} \quad (15)$$

The computation of the IFC index as described by (12) depends on the way the parameters  $Z(\mathbf{i}, k)$  and  $\beta(\mathbf{i}, k)$  are estimated. Thus, (12) can be thought of as a probabilistic IFC index (similar to the probabilistic SSIM index). We obtain the sample IFC index by substituting the estimates from (13) and (14) into (12):

$$\begin{aligned} \text{IFC}[\mathbf{x}(\mathbf{i}, k), \mathbf{y}(\mathbf{i}, k)] &= -\frac{1}{2} \log_2 \left\{ 1 - \left[ \frac{\sigma_{\mathbf{x}(\mathbf{i}, k)\mathbf{y}(\mathbf{i}, k)}}{\sigma_{\mathbf{x}(\mathbf{i}, k)}\sigma_{\mathbf{y}(\mathbf{i}, k)}} \right]^2 \right\} \\ &= -\frac{1}{2} \log_2 \{ 1 - s[\mathbf{x}(\mathbf{i}, k), \mathbf{y}(\mathbf{i}, k)]^2 \} \end{aligned} \quad (16)$$

$C_3 = 0$  in (16). Thus, the IFC index at a location in a sub-band is a *monotonic function* of the structure term of the SSIM index computed at the same location in the same sub-band, as long as the window used in both metrics for estimation purposes are identical. The IFC index is applied in the sub-band filtered domain and is better able to account for the contrast masking properties of the HVS than SSIM and is very closely related to the multi-scale SSIM index in [9]. In fact, if the same frequency decomposition and estimation windows are used in IFC and multi-scale SSIM, the local quality indices obtained using both metrics will be equivalent due to the monotonic relationship described by (16). These assumptions are not very restrictive since the choice of window in [9] and [3] is arbitrary. Further, the development of multi-scale structural similarity is independent of the underlying frequency decomposition, as is the information theoretic framework (the statistical models used for natural scene modeling and distortion modeling are applicable for any choice of frequency decomposition). However, no saturation constant appears in (16) and hence, the IFC index will suffer from instability issues in regions where the signal energy in a sub-band is very low, similar to the UQI index.

### 4.2. The VIF metric

An improved extension of the IFC metric, known as the Visual Information Fidelity (VIF) metric, uses a more sophisticated vector GSM model [15]. However, in this paper, we restrict our analysis to a scalar version of the VIF metric, where the natural scene model is identical to that used in the scalar IFC index. In addition to the gain

and additive noise distortion channel, VIF also models the HVS as a distortion channel that both the reference and distorted images pass through:

$$\begin{aligned} Y(\mathbf{i}, k) &= \beta(\mathbf{i}, k)X(\mathbf{i}, k) + V(\mathbf{i}, k) + W(\mathbf{i}, k) \\ \chi(\mathbf{i}, k) &= X(\mathbf{i}, k) + W(\mathbf{i}, k) \end{aligned}$$

Here,  $W(\mathbf{i}, k)$  is an AWGN field that models the neural noise in the HVS. The variance of this neural noise is assumed to be the same for all sub-bands and we denote it by  $\kappa$ .  $\chi(\mathbf{i}, k)$  is the output of the HVS channel that the reference image passes through. Now, the VIF index is given by:

$$\text{VIF}[X(\mathbf{i}, k), Y(\mathbf{i}, k)] = \frac{I[X(\mathbf{i}, k), Y(\mathbf{i}, k)|Z(\mathbf{i}, k)]}{I[\chi(\mathbf{i}, k), X(\mathbf{i}, k)|Z(\mathbf{i}, k)]} \quad (17)$$

$$= \frac{\frac{1}{2} \log_2 \left( \frac{\beta(\mathbf{i}, k)^2 Z(\mathbf{i}, k)^2 + \sigma_v(k)^2 + \kappa}{\sigma_v(k)^2 + \kappa} \right)}{\frac{1}{2} \log_2 \left( \frac{Z(\mathbf{i}, k)^2 + \kappa}{\kappa} \right)} \quad (18)$$

Estimation details of all parameters of the scalar VIF model are identical to those provided for the IFC model. The one extra parameter in this model, namely the variance of the neural noise  $\kappa$ , is hand optimized in [4] and chosen to be 0.1. As before, (18) represents the probabilistic VIF index. Replacing all quantities in (17) with their estimates, we obtain the sample VIF index as

$$\text{VIF}[x(\mathbf{i}, k), y(\mathbf{i}, k)] = \frac{\log \left[ 1 - s[\mathbf{x}(\mathbf{i}, k), \mathbf{y}(\mathbf{i}, k)]^2 \left( \frac{\sigma_{\mathbf{y}(\mathbf{i}, k)}^2}{\sigma_{\mathbf{y}(\mathbf{i}, k)}^2 + \kappa} \right) \right]}{\log \left[ 1 - \left( \frac{\sigma_{\mathbf{x}(\mathbf{i}, k)}^2}{\sigma_{\mathbf{x}(\mathbf{i}, k)}^2 + \kappa} \right) \right]} \quad (19)$$

We observe that the VIF index includes a structure comparison term ( $s[\mathbf{x}(\mathbf{i}, k), \mathbf{y}(\mathbf{i}, k)]$ ) as well as a contrast comparison term (due to the appearance of functions of  $\sigma_{\mathbf{y}(\mathbf{i}, k)}^2$  and  $\sigma_{\mathbf{x}(\mathbf{i}, k)}^2$  in the numerator and denominator respectively), similar to the SSIM index. One of the properties of the VIF index observed in [4] was the fact that it can predict improvement in quality due to contrast enhancement. Our analysis of the VIF index explains this effect since the correlation coefficient between a contrast enhanced image and the reference image is 1. The VIF index is  $> 1$  in this case since  $\sigma_{\mathbf{y}(\mathbf{i}, k)} > \sigma_{\mathbf{x}(\mathbf{i}, k)}$ . Additionally, the VIF index avoids certain numerical instabilities that occur in the IFC, since the IFC goes to  $\infty$  as  $s[\mathbf{x}(\mathbf{i}, k), \mathbf{y}(\mathbf{i}, k)]$  goes to 1. The form of the contrast comparison in (19) ensures that the VIF index is 1 when the reference and test images are identical.

## 5. CONCLUSIONS

The goal of this paper was to analyze the properties of two recently developed image QA paradigms: the structural similarity framework and the information theoretic framework. We showed that the structure term in SSIM can be interpreted as a divisive normalization model of contrast gain control mechanisms in the HVS, thus establishing a relation between SSIM and HVS based quality metrics. We then studied the information theoretic metrics and proved that the IFC index is equivalent to the structure term of the SSIM index applied between sub-band filtered coefficients. Our results indicate that the current instantiation of the IFC (with a jointly Gaussian assumption on the reference and test image coefficients) provides no advantage over multi-scale SSIM indices locally and further investigation of suitable channel models and their relation to visual quality is needed. We also showed that the VIF index improves over the

IFC due to the use of a contrast comparison function, in addition to the structure comparison. In the future, we would like to extend our analysis to the vector VIF model.

## 6. REFERENCES

- [1] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, 2002.
- [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, 2004.
- [3] H. R. Sheikh, A. C. Bovik, and G. de Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2117–2128, 2005.
- [4] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, 2006.
- [5] A. B. Watson and J. A. Solomon, "Model of visual contrast gain control and pattern masking," *J. Opt. Soc. Am. A*, vol. 14, no. 9, pp. 2379–2391, Sep. 1997.
- [6] P. C. Teo and D. J. Heeger, "Perceptual image distortion," in *Proc. IEEE Int. Conf. Image Proc.*, vol. 2, 1994, pp. 982–986 vol.2.
- [7] 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. IEEE Int. Conf. Acoustics, Speech and Signal Proc.* Glasgow, UK: IEEE, 1989.
- [8] A. R. Reibman and D. Poole, "Characterizing packet-loss impairments in compressed video," in *Proc. IEEE Int. Conf. Image Proc.*, San Antonio, TX, 2007.
- [9] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. Asilomar Conf. Signals, Systems and Computers*, Pacific Grove, CA, 2003.
- [10] Z. Wang and E. P. Simoncelli, "Translation insensitive image similarity in complex wavelet domain," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Proc.*, Philadelphia, PA, 2005.
- [11] S. de Waele and M. J. Verberne, "Coding gain and tuning for parametrized visual quality metrics," in *Proc. IEEE Int. Conf. Image Proc.*, San Antonio, TX, 2007.
- [12] M. K. Mihcak, I. Kozintsev, K. Ramchandran, and P. Moulin, "Low-complexity image denoising based on statistical modeling of wavelet coefficients," *IEEE Signal Process. Lett.*, vol. 6, no. 12, pp. 300–303, 1999.
- [13] E. P. Simoncelli, "Modeling the joint statistics of images in the wavelet domain," in *Proc. SPIE*, vol. 3813, Denver, CO, 1999.
- [14] I. M. Gelfand and A. M. Yaglom, "Calculation of the amount of information about a random function contained in another such function," *Amer. Math. Soc. Transl.*, vol. 12, no. 2, pp. 199–246, 1959.
- [15] M. J. Wainwright and E. P. Simoncelli, "Scale mixtures of gaussians and the statistics of natural images," in *Adv. Neural Info. Proc. Systems*, S. A. Solla, T. Leen, and S.-R. Muller, Eds., vol. 12, 1999, pp. 855–861.