Perceptual quality assessment of color images using adaptive signal representation

Umesh Rajashekar[a], Zhou Wang[b], and Eero P. Simoncelli[a]

[a]Center for Neural Science, and Courant Institute of Mathematical Sciences, New York University, New York, USA

[b]Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Canada

ABSTRACT

Perceptual image distortion measures can play a fundamental role in evaluating and optimizing imaging systems and image processing algorithms. Many existing measures are formulated to represent "just noticeable differences" (JNDs), as measured in psychophysical experiments on human subjects. But some image distortions, such as those arising from small changes in the intensity of the ambient illumination, are far more tolerable to human observers than those that disrupt the spatial structure of intensities and colors. Here, we introduce a framework in which we quantify these perceptual distortions in terms of "just intolerable differences" (JIDs). We first construct a set of spatio-chromatic basis functions to approximate (as a first-order Taylor series) a set of "non-structural" distortions that result from changes in lighting/imaging/viewing conditions. These basis functions are defined on local image patches, and are adaptive, in that they are computed as functions of the undistorted reference image. This set is then augmented with a complete basis arising from a linear approximation of the CIELAB color space. Each basis function is weighted by a scale factor to convert it into units corresponding to JIDs. Each patch of the error image is represented using this weighted overcomplete basis, and the overall distortion metric is computed by summing the squared coefficients over all such (overlapping) patches. We implement an example of this metric, incorporating invariance to small changes in the viewing and lighting conditions, and demonstrate that the resulting distortion values are more consistent with human perception than those produced by CIELAB or S-CIELAB.

Keywords: color image quality assessment, adaptive signal decomposition, CIELAB, S-CIELAB

1. INTRODUCTION

A perceptually accurate measure of color image distortion can benefit many real-world applications. Examples include evaluating the quality of color prints and the effects of gamut mapping in printers, and assessing the performance of algorithms for color image restoration, compression, transmission, white balance correction, and demosaicing. Given a reference color image, $\mathbf{x}$, and its distorted version, $\mathbf{y}$, one of the simplest distortion measures is the mean squared error (MSE) between the two images in the pixel domain: $D_{MSE} = \frac{1}{N}\|\mathbf{x} - \mathbf{y}\|^2$ where N is the number of pixels in the image. For color images, D_{MSE} is often computed independently for the red, green, and blue (RGB) color channels and averaged together to get the final distortion. However, since the RGB space is perceptually non-uniform (i.e. the set of equally distorted RGB vectors is not isotropic around the reference vector, and the geometry of this set is different for different reference vectors), this distortion metric fails to produce distortion values that are consistent with human perception.[1]

There is an extensive literature on the perception of color, from both scientific and engineering perspectives, that describes color spaces that are more compatible with human perception than the RGB color space. At present, the most widely used of these is CIELAB, which is optimized for quantifying perceived color differences of large uniform targets.[2] A distortion measure based on MSE in CIELAB space can perform better than in

Author contact information (address correspondence to U.R.):
U.R.: umesh@cns.nyu.edu
Z.W.: z.wang@ece.uwaterloo.ca
E.P.S: eero.simoncelli@nyu.edu

the original RGB space, but still offers a poor description of human perception, primarily because it does not take into account human sensitivities to spatial structures. The recently-developed spatial CIELAB (S-CIELAB) metric[3] corrects this by incorporating spatial contrast sensitivity weighting into each of the CIELAB color bands, and provides good predictions of human performance in detecting distortions in halftone color images.[1,4] There have been several extensions to these color difference based methods that incorporate models of color appearance (e.g., CIECAM02[5]) culminating in image color appearance models that combine color appearance models with spatial properties of human vision.[6,7] Broadly speaking these techniques attempt to find a function $\mathbf{f}(\cdot)$ that maps the images to a perceptual space, where distortion is then computed as the squared error between the original and distorted images: $D_{perceptual} = \|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})\|^2$. There have also been attempts to extend other spatial-domain error measures by applying them to individual color channels and combining the results. For example, the structural similarity (SSIM) approach[8] has been shown to produce reasonable results when averaged over the channels in transformed color spaces.[9,10]

Measuring distortion using D_{MSE} has the advantage that the error function has a quadratic form, which can be easily optimized while building imaging systems and image processing algorithms. This distortion metric is however known to correlate very poorly with subjective measurements of quality.[11,12] On the other hand, $D_{perceptual}$, was designed to match human perception of color, but has the disadvantage that the function $\mathbf{f}(\cdot)$ is nonlinear and thus, it is difficult to use the metric as an objective function that is to be optimized for a image-processing algorithm. Further, the fact that CIELAB (or S-CIELAB) does not predict subjective measurements of distortions such as JPEG distortions,[1] suggests that there is much scope for improvement in the design of $\mathbf{f}(\cdot)$.

In this paper, we propose a color distortion metric that is motivated by a simple observation that certain image distortions, even when noticeable, are not bothersome or intolerable to human observers.[8] Examples of such distortions include slight changes to the ambient illumination of the scene or slight changes in the viewing position of the observer. We use an locally adaptive spatio-chromatic signal decomposition[13] to separate these "non-structural" image distortions (those that do not affect the perception of the structures of the objects in a visual scene) from the remaining structural distortions. Specifically, for each patch of an image, we represent the error as a linear combination of an overcomplete set of basis functions. The non-structural basis functions are computed as the gradients of various non-structural distortions, rescaled to units of JNDs, thus allowing a linear (first-order Taylor series) approximation of those distortions. The matrix of structural basis functions corresponds to the Jacobian of the transformation from RGB to CIELAB space, and thus allows a linear approximation of the JNDs captured by CIELAB. The error for an individual patch is measured as the weighted sum of squared coefficients in this basis, where the weighting is chosen to convert the units of the non-structural coefficients from JNDs to 'just intolerable differences' (JIDs). The global distortion measure is then the sum of these errors, over the full set of (overlapping) patches. We develop a specific example of this framework using spatio-chromatic adaptive basis functions to capture changes in intensity of the illuminant and a simple geometric distortion (horizontal shift), and demonstrate via simple examples that our distortion measure provides a better match to human perception than CIELAB or S-CIELAB.

2. A GEOMETRIC OVERVIEW OF THE ADAPTIVE BASIS FRAMEWORK

We begin with a geometric overview of the adaptive basis framework for grayscale image quality assessment.[13] It is perhaps easier to appreciate the intuition behind this framework by considering the distortion of an original image $\mathbf{x}$ in two dimensions as shown in Fig. 1. The two axes could be, for example, the grayscale values of two neighboring pixels. An iso-quality contour around a reference vector $\mathbf{x}$ is a set of points in this space that represent distorted versions of $\mathbf{x}$ with the same visual quality. If human vision were well described by mean squared error in the pixel domain, the iso-quality contours for the reference vector $\mathbf{x_1}$, would be circles centered on $\mathbf{x_1}$, as illustrated in the Fig. 1a. If we consider other reference vectors, $\mathbf{x_2}$ or $\mathbf{x_3}$, the iso-quality contours are also circles and have the *exact* same size as the one centered at vector $\mathbf{x_1}$.

It is well known that the human visual system is more sensitive to errors along certain dimensions that others. The weighted mean squared error distortion metric, for example, uses the human contrast sensitivity function to weigh errors in high frequency spatial channels differently than those in low/mid frequencies. Geometrically, the reference vectors are first transformed to the frequency space using the Fourier transform. The iso-quality circles are then stretched *along* the cardinal directions of this space as show in Fig. 1b. The iso-quality contours
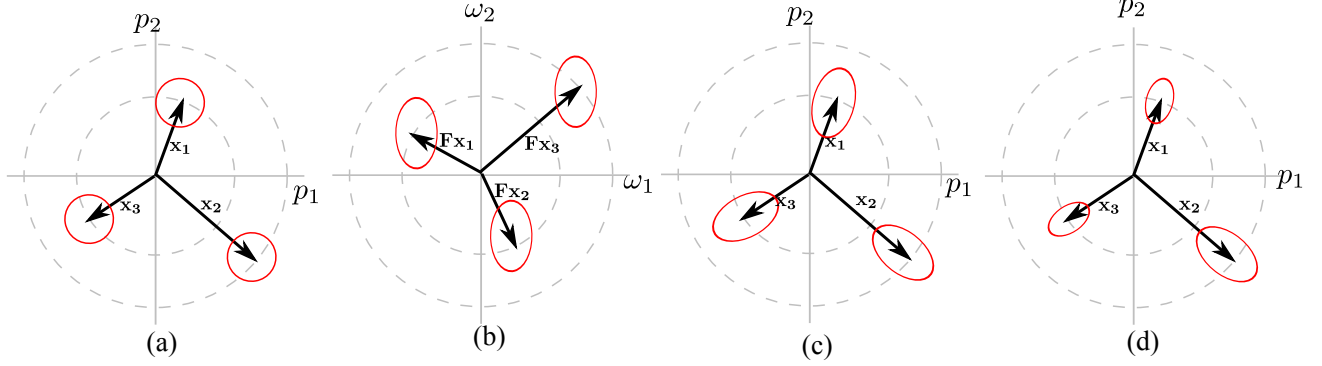
Figure 1. The red circle/ellipses represent contours of equal perceptual quality around a vector $\mathbf{x}$. (a) MSE in the pixel domain corresponds to identical circles for all vectors (b) Weighted mean-squared error in the frequency domain elongates the iso-quality contour along the cardinal axis. $\mathbf{Fx}$ represents the Fourier transform of $\mathbf{x}$. (c) An adaptive basis that is insensitive to modest changes in illumination intensity generates iso-quality contours that are elongated in the direction of the reference vector. (d) An adaptive basis from (c) that modifies the size and shape of the iso-quality ellipse in proportion to the intensity of the reference vector in accordance with Weber's law.

are now ellipses, but as in the case of MSE in the pixel domain, have the same size and shape for all reference vectors.

The fundamental philosophy of the adaptive basis framework is that there are certain directions in the image space that are less sensitive to distortions than others, and that *these directions generally depend on the reference image.* As a simple example, consider scaling the vector $\mathbf{x}$ such that the distorted vector is $\mathbf{y} = a\mathbf{x}$. Such a rescaling would arise from an increase or decrease in the ambient illumination. The human visual system discounts small changes in illumination – even when they are visible, they are far less objectionable than changes that distort the spatial pattern of intensities, such as additive white noise. Thus we can modify an iso-quality contour around $\mathbf{x_1}$ to allow more distortion along the direction of $\mathbf{x_1}$ than in a direction perpendicular to $\mathbf{x_1}$, as indicated by the ellipses in Fig. 1c. An iso-quality contour around $\mathbf{x_2}$ is also elliptical, but with its principal axis aligned with $\mathbf{x_2}$. Clearly, stretching the iso-quality ellipse along any fixed direction (such as Fig. 1b) cannot capture this behavior. Fig. 1d shows a modified version of this framework, where the iso-quality ellipse is not only is aligned with the reference vector, but also changes its shape in proportion to the magnitude (length) of the reference vector in accordance with Weber's law. Figures1c and 1d model an observer's invariance to changes in the intensity of the illuminant. The art of the adaptive basis framework lies in defining similar signal-dependent non-structural directions and the relative scaling of the iso-quality surface along these directions. By doing so, we define an adaptive subspace in which the errors are visually less objectionable. Signal-dependent modification of iso-quality contours can also be found in other quality metrics.[8, 14]

We can quantify the adaptive basis distortion measure using standard linear algebraic methods.[13] Given a reference signal, $\mathbf{x}$ and a distorted version, $\mathbf{y}$, we write the error, $\mathbf{y} - \mathbf{x}$, as a weighted sum over a set of basis functions:

$$\mathbf{y} - \mathbf{x} = \mathbf{L}(\mathbf{x})\mathbf{c} = c_1\mathbf{l_1}(\mathbf{x}) + c_2\mathbf{l_2}(\mathbf{x}) + ... \tag{1}$$

where $\mathbf{L}(\mathbf{x})$ represents a matrix whose columns contain the basis functions $\mathbf{l_k}(\mathbf{x})$. The matrix is divided into two sub-matrices $\mathbf{L}(\mathbf{x}) = [\mathbf{A}(\mathbf{x}) \mid \mathbf{B}]$, where $\mathbf{A}(\mathbf{x})$ represents the collection of adaptive basis vectors that are used to capture the non-structural distortions. These vectors are typically derived from the reference signal $\mathbf{x}$. For example, the adaptive basis that discounts for the illumination change in Fig. 1c and d is simply the vector $\mathbf{x}$. The matrix $\mathbf{B}$ contains a set of basis vectors spanning the space of the input vector and is used to capture structural distortions. For example, $\mathbf{B}$ could be an identity matrix, or the matrix representing the inverse Fourier transform. The distortion measure is now defined as a function of the coefficients, $\mathbf{c}$. In particular,

$$D(\mathbf{x}, \mathbf{y}) = \min_{\mathbf{c}:\mathbf{L}(\mathbf{x})\mathbf{c}=\mathbf{y}-\mathbf{x}} \|\mathbf{Wc}\|^2 \tag{2}$$

where, $\mathbf{W}$ is a diagonal weight matrix, which can be used to define the shape of the iso-quality ellipse. Since $\mathbf{L}(\mathbf{x})$ is overcomplete, solving for the coefficients, $\mathbf{c}$, is a least squares optimization problem. After some algebraic manipulation, the total distortion can be written as

$$D(\mathbf{x}, \mathbf{y}) = \left\| \mathbf{W}^{-1} \mathbf{L}(\mathbf{x})^T \left(\mathbf{L}(\mathbf{x}) \mathbf{W}^{-2} \mathbf{L}(\mathbf{x})^T \right)^{-1} (\mathbf{y} - \mathbf{x}) \right\|^2 . \tag{3}$$

Efficient implementations of (3) are discussed in.[13] In practice, this value is computed "convolutionally" on all local patches in the image, resulting resulting in a distortion map, which is then spatially averaged to produce a single quality measure.

3. QUANTIFYING DISTORTIONS IN COLOR IMAGES

In this section, we describe our choice of the adaptive basis, fixed basis, and the corresponding weight matrix for measuring color distortions.

3.1 Defining the basis vectors

Our framework is based on the hypothesis that human judgments of image quality are relatively insensitive to small changes in the viewing or imaging conditions. For example, a slight decrease or increase in the intensity of the illumination level, even if visible, does not cause a noticeable degradation in the appearance of the image. To capture this variation in the image, one of the adaptive basis vectors is chosen to be the reference vector itself (see Fig. 1c and d). It is also reasonable to assert that slight changes in the observer's viewing position should not alter their perception of image quality. As an example of this, consider a horizontally translated version of the reference image. A first-order Taylor approximation expresses the distorted image as the a linear combination of the reference image and its horizontal derivative. Accordingly, the second adaptive direction was chosen to be the horizontal derivative of the reference image.

Note that the color space of the reference images was not explicitly stated in (1). In an earlier implementation, we demonstrated the use of the adaptive basis framework in the sRGB pixel space.[15] Given the extensive history of modeling distortions via JNDs in the CIELAB space, it is preferable to work in a space where the coefficients, $\mathbf{c_i}$, in (1), and the resulting distortion metric in (2) are represented in units of JND. However, the transformation from sRGB to CIELAB is non-linear. To retain a quadratic error formulation in the original pixel space (for optimization purposes), we approximated distortions in the CIELAB space using the Jacobian matrix of the transformation that maps sRGB to CIELAB. We note that previous work has used signal-dependent basis vectors for analyzing errors, with the basis arising from the Jacobian of either physiological[16] or psychophysical[2,14] models of human vision.

We can now define our basis functions in any space, and then transform them into CIELAB space before measuring the perceptual distortion using (3). For the results shown in this paper, we used the identity basis to capture structural distortions in CIELAB space. Thus, for reference images in the sRGB space, the matrix $\mathbf{B}$ is the Jacobian of the transformation that converts sRGB to CIELAB. Using only this basis leads to (a linear approximation) of the standard CIELAB computation of color difference (Delta E 1976). In our implementation, the distortion measure (3) was computed on patches of size 11×11.

3.2 Learning the weights

The weight matrix, $\mathbf{W}$, in (3) specifies the relative tolerability (measured in JNDs) of errors in the direction of each basis vector. We made two simplifying assumptions to allow us to determine these values through a set of subjective measurements. First, we assume that JIDs are linearly proportional to JNDs. And second, we assume that JIDs can be estimated independently for each basis vector in $\mathbf{L}$ (this cannot be strictly correct: the overcompleteness of $\mathbf{L}$ means that adding distortion along one direction influences coefficients along other directions). To compute the JIDs along the intensity basis, we added distortion along the direction of the reference vector until the distorted image was considered to be barely intolerable. The number of JNDs of distortion that was added to achieve this was then measured in the linearized CIELAB space. The reciprocal of the average value of this JND across several image patches was used as the weight in the matrix $\mathbf{W}$ to convert JNDs to JIDs

| JND = 0 | JND = 6 | JND = 12 | JND = 18 | JND = 24 |

(a) Intensity Noise

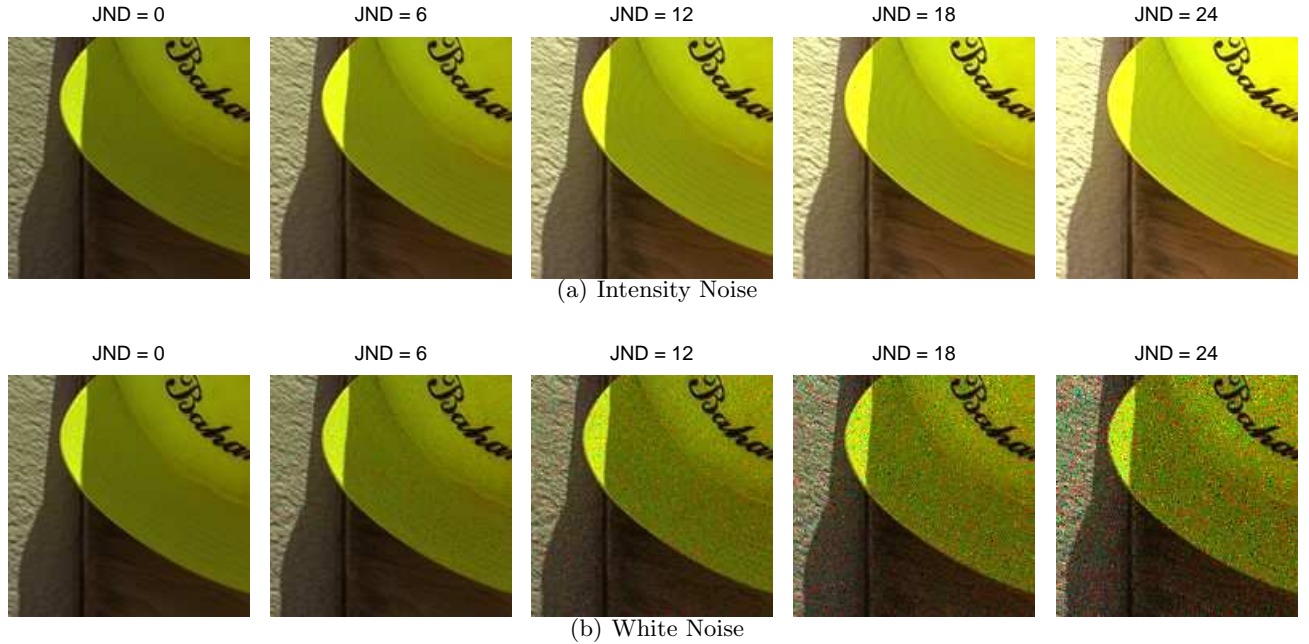| JND = 0 | JND = 6 | JND = 12 | JND = 18 | JND = 24 |

(b) White Noise

Figure 2. Examples of images corrupted by (a)intensity noise (b) white noise. The two images in each column have the same JND distortion as measured by the mean ΔE in the linearized CIELAB space. The intensity distortion is visible, but not intolerable even at a JND of 24. The white noise distortion is not unacceptable until about 18 JNDs for this image.

for the intensity adaptive basis. We repeated this using white noise distortions to get the JND to JID conversion factor for the fixed (identity) basis as well. Figure 2 shows an example of an image patch distorted by intensity and white noise. Each column shows the image patch distorted by the same number of JNDs as measured by ΔE in the linearized CIELAB space. We see from this illustration that even though the intensity distortion is clearly visible, it is tolerable even at 24 JNDs. The white noise distortion becomes intolerable around 18 JNDs for this image. Computing the JND to JID conversion for the shift basis requires a different approach because the JND as measure via the CIELAB space is only valid for color distortions, and not for geometric distortions such as spatial shifts. In a recent study,[17] invariance to geometric and photometric distortions was measured using MSE between the reference and distorted image. The authors conclude that observers are about 8 times less sensitive to horizontal shifts than distortions introduced by white noise. We used their finding and set the weights for the horizontal shift basis to be 1/8th the weight of the white noise basis. Putting this all together, the weights for the intensity, shift, and fixed basis were set to $1/24, 1/96$, and $1/12$ respectively.

4. RESULTS

Figures 3 and 4 show an original image and three different distorted versions of this image: a horizontal shift, an intensity increase or decrease, and additive white noise. The three distorted versions in each image had approximately the same MSE. We then measured the distortion between the original image and each distorted image using three methods: (a) the average of the ΔE map using CIELAB, (b) the average of the ΔE map using S-CIELAB and (c) using the average of the adaptive error D defined by (3). First consider the image in Fig. 3b, where the distorted image is a horizontally shifted version of the original image, corresponding to a slight translation the observer's position. Despite the fact that this distortion is hardly noticeable, the mean distortion reported by the CIELAB (8.69 JNDs) suggests that this is a very noticeable distortion. The spatial blurring introduced by S-CIELAB suppresses the error between the original and the shifted image resulting in a distortion value (4.03 JNDs) that is lower than the one produced by CIELAB. The adaptive basis distortion (0.02 JIDs) on the other hand suggests that the shifted image is almost identical to the original. This is due to

Figure 3. Examples of distortions to the 'Bikes' image. All distorted images have an MSE of approximately 1065. (a) Original: CIELAB = 0.0 JND; S-CIELAB=0.0 JND; D=0.0 JID. (b) Horizontal Shift: CIELAB = 8.69 JNDs; S-CIELAB= 4.03 JNDs; D=0.02 JIDs. (c)Intensity Increase: CIELAB = 8.75 JNDs; S-CIELAB = 8.85 JNDs; D=1.17 JIDs. (d) White Noise: CIELAB=35.40 JNDs;S-CIELAB= 7.97 JNDs; D=24.89 JIDs

Figure 4. Examples of distortions to the 'Church and Capitol' image. All distorted images have an MSE of approximately 930. (a) Original Image: CIELAB = 0.0JND; S-CIELAB=0.0JND; D=0.0JID. (b) Horizontal Shift: CIELAB = 6.18 JNDs; S-CIELAB = 2.81 JNDs; D=0.03 JIDs. (c) Intensity Decrease: CIELAB = 5.96 JNDs; S-CIELAB = 6.07 JNDs; D=0.56 JIDs. White Noise: CIELAB =24.10 JNDs; S-CIELAB = 6.09 JNDs; D=20.72 JIDs

the fact that one of the adaptive directions captures and discounts errors cause by small horizontal shifts. In Fig. 3c the distorted image corresponds to viewing the original image under brighter illumination. Unlike Fig. 3b this distortion is very noticeable, but by no means intolerable. Once again, the CIELAB's distortion measure is high (8.75 JNDs). Since the distortion is a smooth variation across the image, the spatial smoothing introduced by S-CIELAB has no effect on the disortion resulting in a value (8.85 JNDs) similar to CIELAB, while the adaptive basis discounts such changes and produces a much smaller distortion value (1.17 JIDs). Finally, in Fig. 3d, we introduce white noise, which is highly noticeable and disturbing. The distortion values for both CIELAB and our method are large (35.40 JNDs and 25.99 JIDs respectively). Since white noise is decorrelated with the original image, the signal-dependent adaptive basis vectors fail to capture this distortion. Most of the error is therefore represented by the fixed basis, **B**, resulting in a CIELAB-like distortion measure. The smoothing introduced by S-CIELAB tends to blur the noise producing a value of 7.97 JNDs. Figure 4 shows another example with similar behavior with the exception that Fig. 4c shows a decrease in the intensity of the illuminant.

5. CONCLUSIONS

We developed a new framework for color image quality metric by extending the adaptive basis concept[13] and showed that this framework is effective at discounting distortions corresponding to naturally occurring changes in viewing or illumination conditions, while still being able to capture other distortions such as noise artifacts. The examples shown here serve as an initial demonstration of the feasibility of the method, but there is much room for refinement in the choice of basis functions and weights. The fixed linear basis of the model can be constructed from Taylor series approximations of any existing quality metric, in place of the linearized CIELAB used in our examples. The performance of the resulting measure can then be no worse than the original metric (assuming the linear approximations hold). As such, we are currently working to incorporate S-CIELAB as the substrate for our fixed basis.

The adaptive basis is meant to capture image changes that arise from changes in viewing or illumination conditions. In previous work,[13] we conjecture that the human visual system has evolved to discount such changes since they do not correspond to changes in the objects/substances one is looking at. There are many such non-structural variations that can be added to the current framework, by simply incorporating the appropriate image-dependent basis functions into the matrix, $\mathbf{A}(\mathbf{x})$ in (1). Examples include: color or position of light sources and sensor position/orientation. Despite this flexibility, once the basis vectors have been computed from the reference image, the error metric is quadratic, allowing for easy optimization. We expect this will prove invaluable when using this metric as an objective function for improving image processing applications such as denoising or compression. In the event that a desired invariance is difficult to describe analytically, it may also be possible to learn the transformations using pairs of undistorted and distorted versions of images.[18]

ACKNOWLEDGMENTS

REFERENCES

[1] Zhang, X. and Wandell, B. A., "Color image fidelity metrics evaluated using image distortion maps," *Signal Processing* **70**, 201–214 (Nov. 1998).

[2] "Industrial color-difference evaluation, publication CIE 116-95," (1995).

[3] Zhang, X. and Wandell, B. A., "A spatial extension of cielab for digital color image reproduction," *SID Journal* (1997).

[4] Zhang, X., Silverstein, D., Farrell, J., and Wandell, B., "Color image quality metric S-CIELAB and its application on halftone texture visibility," in [*Compcon '97. Proceedings, IEEE*], 44–48 (1997).

[5] "A colour appearance model for colour management applications: Ciecam02 pub. 159," (2004).

[6] Johnson, G. M., *Measuring Images: Differences, Quality and Appearance*, PhD thesis, Rochester Institute of Technology (2003).

[7] Fairchild, M. D. and Johnson, G. M., "icam framework for image appearance, differences, and quality," *Journal of Electronic Imaging* **13**(1), 126–138 (2004).

[8] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P., "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Processing* **13**, 600–612 (April 2004).

[9] Toet, A. and Lucassen, M. P., "A new universal colour image fidelity metric," *Displays* **24**, 197–207 (Dec. 2003).

[10] Bonnier, N., Schmitt, F., Brettel, H., and Berche, S., "Evaluation of spatial gamut mapping algorithms," in [*Proc. of the 14th Color Imaging Conference*], (2006).

[11] Girod, B., "Psychovisual aspects of image processing: What's wrong with mean squared error?," in [*Proc. Seventh Workshop on Multidimensional Signal Processing*], P.2 (Sept. 1991).

[12] Wang, Z. and Bovik, A. C., "Mean squared error: Love it or leave it? a new look at signal fidelity measures," **26**, 98–117 (Jan. 2009).

[13] Wang, Z. and Simoncelli, E. P., "An adaptive linear system framework for image distortion analysis," in [*Proc. IEEE Int. Conf. on Image Processing ICIP 2005*], **3**, III–1160–3 (11–14 Sept. 2005).

[14] Pons, A. M., Malo, J., Artigas, J. M., and Capilla, P., "Image quality metric based on multidimensional contrast perception models," *Displays* **20**, 93–110 (Aug. 1999).

[15] Rajashekar, U., Wang, Z., and Simoncelli, E., "Quantifying color image distortions based on adaptive spatio-chromatic signal decompositions," in [*Proc 16th IEEE Int'l Conf on Image Proc*], IEEE Computer Society, Cairo, Egypt (Nov 7-11 2009).

[16] Portilla, J., "A perceptual bayesian estimation framework and its application to image denoising," in [*EU-SIPCO*], (2006).

[17] Kingdom, F. A. A., Field, D. J., and Olmos, A., "Does spatial invariance result from insensitivity to change?," *J Vis* **7**(14), 11.1–1113 (2007).

[18] Miao, X. and Rao, R. P. N., "Learning the lie groups of visual invariance," *Neural Computation* **19**, 2665–2693 (Oct. 2007).