THE PERCEPTIBILITY OF VIDEO ARTIFACTS:
A PERSPECTIVE FROM COLOR SCIENCE

Mark D. Fairchild

Munsell Color Science Laboratory, Chester F. Carlson Center for Imaging Science,
Rochester Institute of Technology, 54 Lomb Memorial Drive, Rochester, NY 14623-5604 USA
mdf@cis.rit.edu

ABSTRACT

In the world of digital image and video processing, encoding and/or compression errors are often assessed in terms of the signal dimensions (*e.g.*, RGB or YC_BC_R) using quantities such as RMS deviation or signal-to-noise ratio with little or no regard for the radiometric linearity of the quantities or the ultimate appearance to observers. When visual models are utilized it is usually with the aim of predicting visibility thresholds and not the perceived magnitudes of clearly visible artifacts. Color scientists quantify the visibility of color changes using color difference metrics (*e.g.*, CIE ΔE^*_{ab}, $\Delta E94$, and $\Delta E2000$) computed in the CIELAB color space. Such metrics have been largely developed and optimized for specifying the magnitudes of small, supra-threshold, color differences of uniform object colors in illuminated environments. Extensions to both approaches are required to accurately model and predict the perception of differences in images and video. This paper provides an overview of error metrics used in colorimetry and their extension with models of spatial and temporal vision to imaging applications. It also examines some of the important viewing condition variables for images and their surroundings and how they are addressed with color and image appearance models. Lastly, some recent and ongoing research on the perception of image differences and quality is described, including issues in visual equivalence and image content. This is a logical extension of the VPQM 2007 presentation, “A Color Scientist Looks at Video,”[1] which stressed the importance of accuracy in encoding, processing, and display of video content to examine the perceptibility artifacts such as those introduced by compression, noise, or other errors as well as differences that are purposefully introduced into images through enhancement algorithms.

1. DIFFERENCES IN IMAGES AND VIDEO

Researchers in image and video processing often find it necessary to quantify the difference between two images. When the differences are introduced by image degradations such as compression artifacts, chrominance subsampling, or errors in decoding for display, the differences can be considered a form of image quality metric to quantify the visibility and perhaps objectionability of the degradations. Sometimes, however, differences can be desirable such as those introduced by image or color enhancement algorithms and the same metrics can be used to quantify those differences. In most cases it is the perceived difference on a display, or displays, that is of most interest. However it is very rare that perceived differences on meaningful displays are actual measured and discussed.

Instead differences are often expressed in terms of the data representing the image. Such data are usually not directly proportional to perceived image color and actually usually do not even represent the physical image veridically. This is because there are inherent linear and nonlinear relationships between the image data and the luminance of the (typically) three (RGB) display channels. These relationships are not channel independent as the displayed R luminance, for example, often depends on all three RGB channels of image data. Most often, these relationships are not properly characterized so there is little knowledge of the ultimate display colorimetry for given image signals. Even with this knowledge, displays and viewing conditions would have to be properly calibrated and characterized to complete the chain from image data to visual stimulus that can be used to compute meaningful perceived differences.

In addition to accurately describing display appearance prior to measuring differences, there are two types of differences that might be of interest, thresholds and magnitudes. Threshold metrics simply describe whether or not an image difference is perceptible, often in terms of probability of detection or a just-noticeable-difference (JND) criterion. Magnitude metrics describe the perceived difference of images with differences clearly above threshold. Sometimes, units of multiple JNDS (*e.g.*, two images differ by 42 JNDs) are used to erroneously describe difference magnitudes. It is well accepted in perceptual science that JNDs do not scale linearly up to perceptual magnitudes.[2,3]

2. IMAGE/VIDEO PROCESSING METRICS

Image-based metrics are normally computed on the simple image data such as RGB or YC_BC_R and have a tenuous, at best, relationship to perceived differences. Such metrics include RMS or mean-squared error (RMSE or MSE), peak signal-to-noise ration (PSNR), and the structural similarity index (SSIM).

MSE is simply the difference between each image element averaged across the entire image (one channel at a time).[4]

$$MSE = \frac{1}{N}\sum_{i}^{N}\left|\hat{x}_i - x_i\right|^2$$

RMSE is the square root of MSE and puts the error metric back into the same units as the image data.

PSNR is the ratio of the maximum possible image data value to the MSE expressed in decibel units.[4,5]

$$PSNR = 10\log_{10}\left(\frac{\max(x_i)^2}{MSE}\right)$$

SSIM is sometimes described as being perceptually based, however examination of the formula illustrates that it has no relation to perception.[6]

$$SSIM(\hat{x},x) = \frac{(2\mu_{\hat{x}}\mu_x + c_1)(2\operatorname{cov}_{\hat{x}x} + c_2)}{(\mu_{\hat{x}}^2 + \mu_x^2 + c_1)(\sigma_{\hat{x}}^2 + \sigma_x^2 + c_2)}$$

These metrics are not capable of describing perceived image difference mainly because they are not applied on perceptual dimensions such as lightness, chroma, and hue. Instead they are applied on just luminance signals under the assumption that all relevant image quality differences are in luminance only, in RGB or nonlinear RGB with some simple strategy for summing the quantities across the three channels if a single metric is desired, or via a similar approach in linear or nonlinear YC_BC_R.

3. COLORIMETRIC DIFFERENCE METRICS

In the world of color science and measurement, differences are measured as magnitudes on scales designed to estimate the perceptual dimensions of lightness, chroma, and hue (and sometimes others). This approach originated with the perceptual description of color space and perceptual scales by Munsell.[7] Other researchers such as Wright and MacAdam approached the same problem from the direction of color thresholds.[8] These research paths culminated in the creation of the CIELAB and CIELUV color spaces and difference formulae in 1976.[9]

Since 1976, research in color difference perception and tolerance specification has established the superiority of CIELAB over CIELUV and focussed on the creation of weighted color difference equations within the CIELAB color space. The most widely used and best performing of these are the CIE94 and CIEDE2000 color difference equations. Once images are expressed in calibrated CIELAB units of L* (lightness correlate), a* (redness-greenness), and b* (yellowness-blueness) or L* (lightness), C^*_{ab} (chroma) and h_{ab} (hue), then the simple CIELAB 1976 color difference is defined as the Euclidean distance between the two colors.

$$\Delta E^*_{ab} = \sqrt{\left(\Delta L^*\right)^2 + \left(\Delta a^*\right)^2 + \left(\Delta b^*\right)^2} = \sqrt{\left(\Delta L^*\right)^2 + \left(\Delta C^*_{ab}\right)^2 + \left(\Delta H^*_{ab}\right)^2}$$

For images, these differences are often averaged across the entire image or other statistics such as a histogram of differences, or certain percentiles are evaluated. Similar approaches can be taken with the more perceptually accurate CIE94 and CIEDE2000 formulae illustrated partially below.

$$\Delta E^*_{94} = \sqrt{\left(\frac{\Delta L^*}{K_L}\right)^2 + \left(\frac{\Delta C^*_{ab}}{1 + K_1 C^*_{ab}}\right)^2 + \left(\frac{\Delta H^*_{ab}}{1 + K_2 C^*_{ab}}\right)^2}$$

$$\Delta E^*_{00} = \sqrt{\left(\frac{\Delta L'}{S_L}\right)^2 + \left(\frac{\Delta C'}{S_C}\right)^2 + \left(\frac{\Delta H'}{S_H}\right)^2 + \left(R_T\frac{\Delta C'\Delta H'}{S_C S_H}\right)}$$

These weighted equations express the differences in terms of lightness, chroma, and hue and then adjust the relative weighting of the difference dimensions depending on the location in color space. Note that there is not space to fully express the derivation or computation of these difference equations in this paper. See [10] for details. It is likely that the CIEDE2000 equation is more complex than required for imaging applications, but the CIE94 equation probably represents a significant advance over a simple CIELAB 1976 difference.

4. VISUAL THRESHOLD MODELS

There has been significant research on video quality and video quality metrics, often aimed at the creation and optimization of encoding/compression/decoding algorithms such as MPEG2 and MPEG4. By analogy, the still-image visible differences predictor of Daly[11] is quite applicable to the prediction of the visibility of artifacts introduced into still images by JPEG image compression. The Daly model was designed to predict the probability of detecting an artifact (*i.e.*, is the artifact above the visual threshold). The CVDM metric[12] represented an extension of the Daly VDP to include all three dimensions of color. Other metrics have been published to examine the probability of detection of artifacts in video (*i.e.*, threshold metrics). Two well-known video image quality models, the Sarnoff JND model and the NASA DVQ model, are briefly described below to contrast their capabilities with models aimed at predicting image difference magnitudes and appearance.

The Sarnoff JND model is the basis of the JNDmetrix software package <www.jndmetrix.com> and related video quality hardware. The model is briefly described in a technical report published by Sarnoff[13] and more fully disclosed in other publications.[14] It is based on the multi-scale model of spatial vision published by Lubin[15,16] with some extensions for color processing and temporal variation. The Lubin model is similar in nature to the Daly model in that it is designed to predict the probability of detection of artifacts in images. These are threshold changes in images often referred to as just-noticeable differences, or JNDs. The Sarnoff JND model has no mechanisms of chromatic or luminance adaptation. The input to the Sarnoff model must first be normalized (which can be considered a very rudimentary form of adaptation). The temporal aspects of the Sarnoff model are also not aimed at predicting the appearance of video sequences, but rather at predicting the detectability of temporal artifacts. As such, the model only uses two frames (four fields) in its temporal processing. Thus, while it is capable of predicting the perceptibility of relatively high frequency temporal variation in the video (flicker) it cannot predict the visibility of low frequency variations that would require an appearance-oriented, rather than JND-oriented, model. While it is well-accepted in the vision science literature that JND predictions are not linearly related to suprathreshold appearance differences, it is certainly possible to use a JND model to try to predict suprathreshold image differences and the Sarnoff JND model has been applied with some success to such data.

A similar model, the DVQ (Digital Video Quality) metric has been published by Watson[17] and Watson *et al.* [18] of NASA. The DVQ metric is similar in concept to the Sarnoff JND model, but significantly different in implementation. Its spatial decomposition is based on the coefficients of a discrete cosine transformation (DCT) making it amenable to hardware implementation and likely making it particularly good at detecting artifacts introduced by DCT-based video compression algorithms. It also has a more robust temporal filter that should be capable of predicting a wider array of temporal artifacts. Like the Sarnoff model, the DVQ metric is aimed at predicting the probability of detection of threshold image differences. The DVQ model also includes no explicit appearance processing through spatial or temporal adaptation, or correlates of appearance attributes.

5. SPATIAL COLOR DIFFERENCE MODELS

Another approach to incorporating the properties of spatial vision into image difference metrics is to combine spatial filtering of the images with the computation of traditional colorimetric difference metrics. An early example of this process is the S-CIELAB model of Zhang and Wandell.[19] Johnson and Fairchild[20] extended the S-CIELAB framework to include more robust spatial filtering techniques and the CIEDE2000 color difference equation.

A more in-depth and theoretical approach was derived by Johnson *et al.* and referred to as a modular image difference metric.[21] This metric included a comparison between two images using first a set of three two-dimensional contrast sensitivity functions for the opponent-colors dimensions (light-dark, red-green, yellow-blue). This was followed by a spatial localization process that increased the importance of differences near edges in the scene, something observers do as well. The next step was local contrast detection to modulate the predicted differences based on the magnitude and direction of a pixel's contrast with respect to its local background. The filtered images were then transformed into uniform color space (typically the IPT space) and from their a map of traditional color difference components was produced and summarized with a variety of statistics.

Johnson's modular image difference metric evolved into the full iCAM image appearance model.[21,22] The iCAM framework has been successfully applied to various image quality predictions such as changes in sharpness and contrast as well as used to render high-dynamic-range still and video images. The framework continues to be a topic of research.

6. TEMPORAL DIMENSIONS

The S-CIELAB approach to combining spatial filtering and the CIELAB color difference metric has also been extended into the temporal domain for application to video difference issues. Two such approaches are ST-CIELAB and SV-CIELAB.

ST-CIELAB[23] utilized two-dimensional spatio-temporal contrast sensitivity functions applied to luminance and chromatic dimensions prior to CIELAB color difference computations. The computed difference were then pooled spatially and temporally to provide the ST-CIELAB difference rating. It should be noted that in the spatial domain the ST-CIELAB filters are one-dimensional rather than the two spatial dimensions represented in the modular image difference metric of Johnson.[21]

Recently, Hirai *et al.* introduced the SV-CIELAB metric.[24] This novel metric uses filtering in the spatial and velocity (rather than temporal frequency) domains. The initial version of SV-CIELAB works only in the luminance dimension (*i.e.* grayscale videos), but the concept could be readily extended to all three color dimensions. Original and distorted video sequences are first converted to luminance, Y. Then the velocity of motion at each pixel location is computed. The image sequence is then filtered using the SV-CSF and CIELAB differences (in this case just L* differences) are computed between the filtered image sequences. Results of a psychophysical experiment illustrated that the SV-CIELAB metric performed significantly better than ST-CIELAB, S-CIELAB, CIELAB alone, PSNR in CIELAB, and SSIM.[24] Interestingly, SSIM performed reasonably well since the image set was limited to one dimension and a relatively small sampling of videos.

The iCAM framework has also been applied to video rendering via temporal adaptation to illustrate the change in appearance of video sequences over time. However it has not been implemented and tested with spatio-temporal or spatio-velocity filtering; something that remains for future research.

7. FUTURE DIRECTIONS & CONCLUSION

It is clear that, while much has been accomplished in the domain of video quality and difference metrics, much remains to be understood, modeled, and tested in realistic viewing situations. One thing is certain however, no significant improvement in predictions can be made without proper and accurate colorimetric calibration and characterization of video displays and use of perceptual correlates, rather than image data dimensions, as the basis for difference and quality metrics. Other interesting and important problems include adaptation to complex environments, different types of filtering techniques, combination of color difference and color appearance models, application to high-dynamic-range and wide-color-gamut display systems, dependency on image content, and a rather new concept of visual equivalence.

The human visual system is very complex in how it adapts to the viewing environment. For example there are well documented adaptation phenomena for color, space, time, spatial frequency, and temporal frequency. There is also a less-well-documented, but very real, adaptation to noise in images.[25] The iCAM framework and its extensions have been applied to many of these dimensions of adaptation (mainly color and space) to predict image appearance and differences.[21,22] Future work will have to aim to incorporate the other adaptation dimensions.

At the most recent IS&T/SID Color Imaging Conference, it was quite clear that the worlds of color science and image quality could benefit from more cross-pollination. In addition to the SV-CIELAB model that was introduced there,[24] there were interesting papers on the use of an adaptive bilateral filter for predicting color image difference[26] and the application of image quality metrics to color gamut mapping.[27] The bilateral filtering technique was a simple approach that effectively combined traditional spatial filtering with the local adaptation metric of Johnson *et al.*[21]

There is also significant progress yet to be made in the area of traditional color difference equations. The most likely advances will be the combination of CIEDE94-like weighted (but not too complex) color difference equations with the CIECAM02 color appearance model.[22] Berns and Xue[28] have recently reported promising results with more certain to come in the future.

Recent video displays are pushing historical boundaries in terms of color gamut volume and dynamic range.[29] It is very likely that image color difference metrics that applied to historical gamuts and dynamic ranges could fail when applied to high-dynamic-range and wide-color-gamut image displays.

Perceived image quality always depends significantly on image content. A review of eye-tracking research with respect to image quality assessment suggests that observers choose just one image area to attend to for any given task and look at no other parts of the image.[30] However, different observers will choose different image areas and this might well be the source of significant inter-observer variability and image dependency in image quality and image difference experiments. Research is being planned to further probe and address this issue.

One final concept that could be of great use in image and video quality assessment is that of *visual equivalence*.[31,32] Images can be considered visually equivalent when either the pixel data result in displays that are imperceptibly different on a pixel-by-pixel basis or when the subject matter is rendered in such a way that the objects in the scene look appropriate even when they might be physically very different from the original. For example when a scene is rendered using computer graphics, great pains can be taken to assure that every reflection and specular highlight in the image is a perfect physical match to the optics of the modeled scene. Alternatively approximations could be made that produce reflected highlights and scene elements that only approximate physical reality while appearing completely plausible and not affecting the perception of material properties. A pixel-by-pixel comparison of such images will show very large differences that observers tend to overlook unless they are specifically brought to their attention. However a model of visual equivalence (*i.e.*, all the objects in the scene look right) would make a prediction that matches human observation. Clearly deriving such a model is just one of the many challenges in the field of image and video quality assessment.

9. REFERENCES

[1] M.D. Fairchild, “A color scientist looks at video,” *3rd International Workshop on Video Processing and Quality Metrics (VPQM),* Scottsdale, Invited Paper 1 (2007).

[2] G.A.Gescheider, *Psychophysics: Method, Theory, and Application, 2nd Ed.*, Lawrence Erlbaum, Hillsdale (1985).

[3] P.G. Engeldrum, *Psychometric Scaling*, Imcotek, Winchester (2000).

[4] H.R. Wu and K.R. Rao, Eds., *Digital Video Image Quality and Perceptual Coding*, CRC Press, Boca Raton (2006).

[5] Q. Huynh-Thu and M. Ghanbari, “Scope of validity of PSNR in image/video quality assessment,” *IEEE Electronics Letters* **44**, 800-801 (2008).

[6] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing* **13**, 600-612 (2004).

[7] A.H. Munsell, *A Color Notation*, Munsell Color Company, Baltimore (1905).

[8] G. Wyszecki and W.S. Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulae, 2nd Ed.*, Wiley, New York (1982).

[9] A.R. Robertson, "The CIE 1976 color-difference formulae," *Color Research and Application* B, 7-11 (1977).

[10] E. Reinhard, E.A. Khan, A.O. Akyüz, and G.M. Johnson, *Color Imaging: Fundamentals and Applications*, AK Peters, Wellesley (2008).

[11] S. Daly, "The Visible Differences Predictor: An algorithm for the assessment of image fidelity," in *Digital Images and Human Vision*, A. Watson, Ed., MIT, Cambridge, 179-206 (1993).

[12] E.W. Jin, X.F. Feng and J. Newell, "The development of a color visual difference model (CVDM)," *Proc. IS&T PICS Conference*, Portland, 154-158 (1998).

[13] Sarnoff Corporation, *JND: A human vision system model for objective picture quality measurement*, Sarnoff Technical Report from www.jndmetrix.com, (2001).

[14] ATIS, O*bjective perceptual video quality measurement using a JND-based full reference technique*, Alliance for Telecommunications Industry Solutions Technical Report T1.TR.PP.75-2001, (2001).

[15] J. Lubin, "The use of psychophysical data and models in the analysis of display system performance," in *Digital Images and Human Vision*, A. Watson, Ed., MIT, Cambridge, 163-178 (1993).

[16] J. Lubin, "A visual discrimination model for imaging system design and evaluation," in *Vision Models for target Detection and Recognition*, E. Peli, Ed., World Scientific, Singapore, 245-283 (1995).

[17] A.B. Watson, "Toward a perceptual video quality metric," *Human Vision and Electronic Imaging III,* SPIE Vol. **3299**, 139-147 (1998).

[18] A.B. Watson, J. Hu, and J.F. McGowan, "DVQ: A digital video quality metric based on human vision," *Journal of Electronic Imaging* **10**, 20-29 (2001).

[19] X. M. Zhang and B. A. Wandell, "A spatial extension to CIELAB for digital color image reproduction," *Proceedings of the SID Symposiums*, 731-734 (1996).

[20] G.M. Johnson and M.D. Fairchild, "A top down description of S-CIELAB and CIEDE2000," *Color Research and Application* **28**, 425-435 (2003).

[21] M.D. Fairchild and G.M. Johnson, "The iCAM framework for image appearance, differences, and quality," *Journal of Electronic Imaging* **13**, 126-138 (2004).

[22] M.D. Fairchild, *Color Appearance Models, 2nd Ed.*, Wiley-IS&T Series in Imaging Science and Technology, Chichester, UK (2005).

[23] X.Tong, D. Heeger and C.B. Lambrecht "Video Quality Evaluation using ST-CIELAB", Proc. SPIE •Vol. **3644**, 185-196 (1999).

[24] K. Hirai, J. Tumurtogoo, A. Kikuchi, T. Nakaguchi, N. Tsumura and Y. Miyake, SV-CIELAB: "Video quality assessment using spatio-velocity contrast sensitivity function," *IS&T/SID 17th Color Imaging Conference*, Albuquerque, 35-41 (2009).

[25] M.D. Fairchild and G.M. Johnson, "Measurement and modeling of adaptation to noise in images," *Journal of the Society of Information Display* **15**, 639-647 (2007).

[26] Z. Wang and J.Y. Hardeberg, "An adaptive bilateral filter for predicting color image difference," *IS&T/SID 17th Color Imaging Conference*, Albuquerque, 28-31 (2009).

[27] Z. Baranczuk, P. Zolliker and J. Giesen, "Image quality measures for evaluating gamut mapping," *IS&T/SID 17th Color Imaging Conference*, Albuquerque, 21-26 (2009).

[28] R. S. Berns and Y. Xue, "Optimizing color-difference equations and uniform color spaces for industrial tolerancing," *Proc. AIC Midterm Meeting*, Color Association of China, 24-28 (2007).

[29] M.D. Fairchild, "High, wide, & deep: Displayed image color appearance and perception," *SID International Symposium*, Los Angeles, 780-782 (2008).

[30] S. Farnand, "From Buswell to Buswell revisited and beyond: Looking at how people look at pictures; eye movements, visual attention, and image saliency," *Unpublished Technical Report* (2009).

[31] G. Ramanarayanan, J.A.Ferwerda, B.J. Walter and K. Bala, "Visual equivalence: Towards a new standard for image fidelity," *ACM Transactions on Graphics* **26**(3), *(SIGGRAPH '07)*, 1-11 (2007).

[32] J.A. Ferwerda, G. Ramanarayanan, K. Bala and B.J. Walter, "Visual equivalence: An object-based approach to image quality," *Proceedings IS&T/SID 16th Color Imaging Conference*, Portland 347-354 (2008).