The Development of A Color Visual Difference Model (CVDM)

Elaine W. Jin[+], *Xiao-Fan Feng*[*] *and John Newell*
Imaging Sciences Engineering & Technology, Xerox Corp.
Webster, New York

Abstract

A color visual difference model (CVDM) was developed to predict the image quality difference of two images. The model is an extension of the two previously published models: The monochromatic Visible Difference Predictor (VDP) by Scott Daly and the color Spatial-CIELAB (S-CIELAB) model by Zhang *et al*.

The CVDM consists of color space conversion, modulation by contrast sensitivity functions, visual masking effect, multiresolution detection mechanisms, and visible color difference calculation. Inputs to the model are a reference image and a processed image, as well as calibration parameters such as viewing distance, resolution of the images, and white point. The output of the model is a CIELAB ΔE map, on which the bright colors represent large ΔE values, and the dark colors represent small visible ΔE values. The model was applied to detect the visibility of blur, noise, grating, and compression artifacts. The results show a better agreement with visual impression than does the S-CIELAB model.

Introduction

A document reproduction process usually consists of multiple steps: the image input terminal (e.g., a scanner), the image processing system, and the image output terminal (e.g., a printer). In many steps of digital processing, it is possible that the output image is not exactly the same as the input one for bandwidth and storage reasons (i.e., image compression). The goal of this type of image processing, in terms of image quality, is to produce an image that looks exactly like the original image.

Two kinds of differences may be used to evaluate image quality. The first is the physical difference, i.e. the difference in features of the images that can be measured physically. For example, the feature may be the spectral power distribution or the temporal intensity variation at any image location. The feature may also be the spatial Fourier component of the image. The measurement of the physical difference is constrained by the precision of the physical device used.

Another is the visual difference. It is well known that human vision has limitations. The color appearance of any light stimulus can be determined by only three independent variables, while its physical counterpart, the spectral power distribution, can be measured in many dimensions (e.g., 401 dimensions in 380 - 780 nm, with 1 nm interval). The fine details that human eyes can discern (also called visual acuity) are lower than 60 cpd (or about 1 sec of visual angle). In the temporal domain the human eye cannot differentiate any variation beyond the flicker fusion frequency of about 50 Hz.

Because human vision has limitations, the physical difference is different from the visual difference. In many cases, the highly precise physical measurement will contain details that the human eyes cannot perceive. If our goal is to make the two images look alike, any precision that goes beyond the limitation of the human eyes will be unnecessary. This is the basis for all the lossy compression where the unnecessary components of the image are discarded to save image size. But how much physical degradation is acceptable to yield visually lossless images? A visual difference model is needed to answer the question.

The visual difference model can be used to predict the visible difference of two images. It usually consists of several steps with each step imitating one step of the human visual system signal processing. The model takes two images as its inputs. One is a reference image, which is the original, undistorted image. Another is a test image, which is usually a distorted image obtained after the reference image goes through some image processing procedures, such as compression/decompression, halftoning, etc. Other input parameters may include the viewing distance and the spatial resolution of the image in dots per inch (dpi) for spatial calibration, and the white point for color appearance calculation. The output of the model is a visual difference map of the two images.

Several visual difference models have been developed and applied in image quality evaluation. Among those models, the Visible Difference Predictor (VDP) model has gained wide attention due to its comprehensiveness and its successful application.[1] As far as color image is concerned, the only published visual difference model is the Spatial-CIELAB (S-CIELAB) model developed at Stanford University.[2] Because our color visual difference model (CVDM) incorporates different components from the two existing visual difference models mentioned above, the two models are reviewed in the following section.

Review of the Two Existing Visual Difference Models

Visible Difference Predictor (VDP)

The VDP model is developed for black and white images and is composed of three steps. The first step is called the amplitude nonlinearity, which focuses on the signal processing at the retinal level. The response of the photoreceptors to the incoming light is a nonlinear process. When the light intensity is low, the photoreceptors may not respond to the light until some specific light level is achieved. This light level is called the absolute threshold. On the other hand, when the light level is extremely high, the response of the photoreceptors may reach a limit and no longer increase with the increase in light level. This is called saturation. The thresholding and saturation of the photoreceptors on the retina are characterized in the amplitude nonlinearity step.

The second step of the VDP model is the spatial filtering using a 2-D contrast sensitivity function, which is a postreceptoral process. A contrast sensitivity function specifies the ability of the human visual system to detect intensity modulation as a function of spatial frequency. Previous studies have shown that the sensitivity of the human visual system varies with spatial frequency. At both the low (below 1 cpd) and high spatial frequencies (beyond 8 cpd) the response of the visual system declines. The response of the visual system reaches its peak at the intermediate frequency range (2-4 cpd). The 2-D contrast sensitivity function simulates this bandpass nature of the human visual system.

The detection mechanism is included as the third step in the VDP model. This step takes place at the cortical level, where the response of the neurons is specific to spatial frequency, orientation, and spatial location. The basic goal of the VDP model is to detect any visible distortion of the test image compared to the reference image. According to the findings of the psychophysical studies, detection should be considered as a multi-channel mechanism instead of a single-channel mechanism. The input image is separated into multiple-channel representations (also called subband images) with varying spatial frequencies and orientations. A visual difference is calculated for each subband image. Then all the visual differences are summed up to form a visible difference map between the test image and the reference image. A probability of detection can be derived from the visible difference.

Spatial-CIELAB Model (S-CIELAB)

The S-CIELAB model is based on numerous psychological studies at Stanford University.[3,4,5,6] Their research proved that by first approximation color and pattern are separable. Based on the data of asymmetric color matching and the assumption of color-pattern separability, a set of spectral response curves for three signal-processing channels is specified. The spatial tuning curves (also called contrast sensitivity functions) for the three channels are also specified. Because the three spectral curves are similar to the opponent-color channels found in other studies, the three channels are labeled the luminance channel (O1), the red/green channel (O2), and the blue/yellow channel (O3). In the S-CIELAB model, the color conversion does not depend on the image spatial pattern, and the spatial tuning does not depend on the image color. This is the so called color-pattern separability.

The S-CIELAB model is a spatial extension to the CIELAB calculation and can be used for measuring color reproduction errors of digital images. The inputs to the model are a reference image and a test image, as well as sample per degree (spatial resolution), white point (viewing condition), and image color space (for color conversion). Each of the two images is first separated into three opponent-color images (O1, O2, & O3 images). All three images go through a spatial-filtering process by convolving with kernels of different sizes and shapes. The filtered images are converted back to the standard CIE color space. The CIELAB formulas are used at the final step to calculate the appearance difference of the two images. For uniform color patches, the result of S-CIELAB is the same as that of the CIELAB. For complex color patterns, the S-CEILAB model predicts the visual difference more accurately than the CIELAB calculation. This model has been used for evaluating various processing methods.[7,8]

Two important signal processing steps are missing in the S-CIELAB model. The first step is the visual masking effect, which refers to a situation in which the threshold for detecting a stimulus is elevated when a masker is present. An example of visual masking is image quality evaluation. When we try to determine whether an image has been distorted or not, we are actually detecting the distortion in the presence of the image contents. If the image is very busy and contains many fine details, the detection of the distortion is harder than when a uniform background is present. In this case the masker is the image itself and the stimulus is the distortion to be detected. The masking effect is both spatial-frequency selective and orientation selective. When the masker deviates in spatial frequency or orientation from the test stimulus, the masking effect becomes weaker and disappears beyond a certain range.

The second missing step is the multiresolution representation.[9] It has been revealed by the previous studies that the cortical representation of the input image is a group of images in separate subbands. Each subband image contains a component of the original image that falls in a specific spatial frequency and orientation range. In a detection task, each subband does its own detection independently. The results from all the subbands are summed up to form an overall detection result.

The two steps mentioned above are actually closely related because the visual masking effect is believed to occur only within each subband. The effect of the visual masking is to elevate the threshold of detection in each subband and hence decrease the overall probability of detection. The visual masking effect and the multiresolution representation are significant factors of the human vision detection and hence should be included in a visual difference model.

An additional part of the S-CIELAB model that needs improvement is the contrast sensitivity function. Previous studies have shown that the luminance contrast sensitivity function should have a bandpass shape while the S-CIELAB uses a lowpass luminance contrast sensitivity function. A bandpass CSF would ignore the mean light level and focus only on spatial modulation. This is actually what happens in the human visual system.[10]

In summary, the VDP model is a comprehensive model in the sense that it simulates visual mechanisms starting at the retinal level and going all the way up to the primary visual cortex. However, the model does not include the color signal processing mechanisms and hence is limited to applications in the monochromatic domain. On the other hand, the S-CIELAB model offers a color space for the representation of chromatic signals in human visual system, as well as a method of output (ΔE) for easy comparison with results of current color appearance models. The goal of the present study is to develop a comprehensive color visual difference model to predict the image quality of digital color images. The model is an extension of the VDP model and the S-CIELAB model in the sense that it will adopt useful steps from both the VDP model and the S-CIELAB model.

Description of the Color Difference Model

The CVDM consists of four steps: color space conversion, contrast sensitivity functions, visual masking effect, and multiresolution detection mechanisms.

Color space Conversion

The first step of the CVDM is to choose a proper color space to represent the color image for subsequent processing. It is natural to use an opponent-color space because many studies have shown that the postreceptoral chromatic signals are expressed in some kind of opponent-color space. The opponent-color space used in the S-CIELAB model was chosen as the color space in the CVDM. It has been shown that this color space is similar to the other previously developed opponent-color spaces.[3]

Contrast Sensitivity Functions

The contrast sensitivity functions (CSFs) used in the CVDM are different from the CSFs used in the S-CIELAB model. The O1 channel, the luminance channel, used the CSF of the VDP model.[1] The curve has a peak spatial frequency of 3 cpd. On both sides of the peak the curve declines and reaches zero at both zero cpd and the cut-off frequency (about 30 cpd). The CSFs for the O2 and O3 channels were derived from a classical study by Mullen.[11] The two CSFs have a lowpass shape. Both have a constant value (normalized to unity) in the low spatial frequency range and decline passing a transition point. The O2 (red/green) CSF has a transition point of 0.8 cpd and a cut-off frequency of 11.5 cpd. The O3 (blue/yellow) CSF has the same transition point, but a lower cut-off frequency (11 cpd). The declining of the two curves follows a straight line on a semi-logarithmic scale. Figure 1 shows the three contrast sensitivity functions used in the O1, O2 and O3 channels.

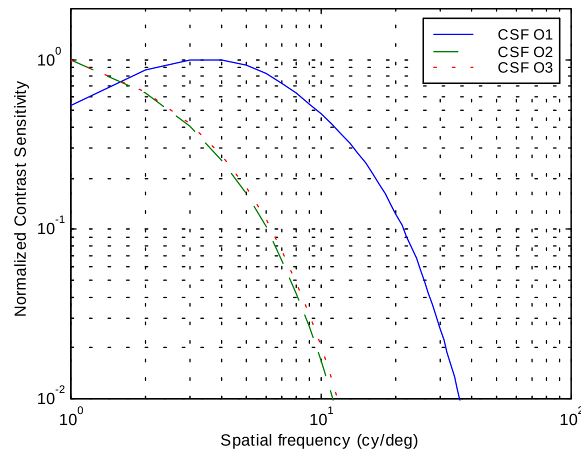

Figure1: The contrast sensitivity functions used in the CVDM.

Frequency Hierarchy in the Three Color Channels

A decomposition of the original images is made before the visual masking effect is considered. This is because the visual masking effect only occurs when the stimulus (distortion to be detected) and the masker (the image itself) are within the same cortical subband, which is specific in spatial frequency, orientation, and color. The numbers of subbands are different for different opponent-color channels. This is because the information contained in the three channels contributes differently to the visual system. Luminance information is considered more important than chromatic information and hence uses the most number of subbands (= 21). The information in the yellow/blue channel (= 9) has the least contribution to the total and hence is processed using less subbands than that is used in the red/green channel (= 13). The overall consideration when choosing subbands is a trade-off between including enough details and computational efficiency.

The Flowchart of the CVDM

The flowchart of the CVDM is shown in Figure 2. Both the reference image (image1) and the test image (image2) are converted to three opponent-color images, O1, O2 and O3 using a 3 by 3 color conversion matrix. The three images are then transformed into the Fourier space. Three contrast sensitivity functions are used to modulate the three spectra, forming a net set of spectra. Each image is then separated into a number of subband images, with each subband having its own optimal spatial-frequency and orientation ranges. In each of the subbands, the visual masking factor (MFi) is determined based on the image content in that band. Then the subband reference and test images are transformed back to the intensity domain (still in the opponent-color space) for calculation of the CSF weighted difference (Δi) between the two images. This CSF weighted difference is divided by the masking factor to form the visual difference in that subband (Δi/MFi). After all subband images are processed, the overall visual difference between the two images is calculated from the subband visual differences by probability summation. This overall visual difference is added to the filtered reference image (image 1') to form a new test image (image2'). The two images are then converted back to a standard color space and the CIELAB formulas are used to calculate the color appearance difference. The key improvement of the CVDM over the S-CIELAB model is the incorporation of masking in the visual difference calculation. This reflects our belief that masking is an important factor in the visual detection of image distortion.

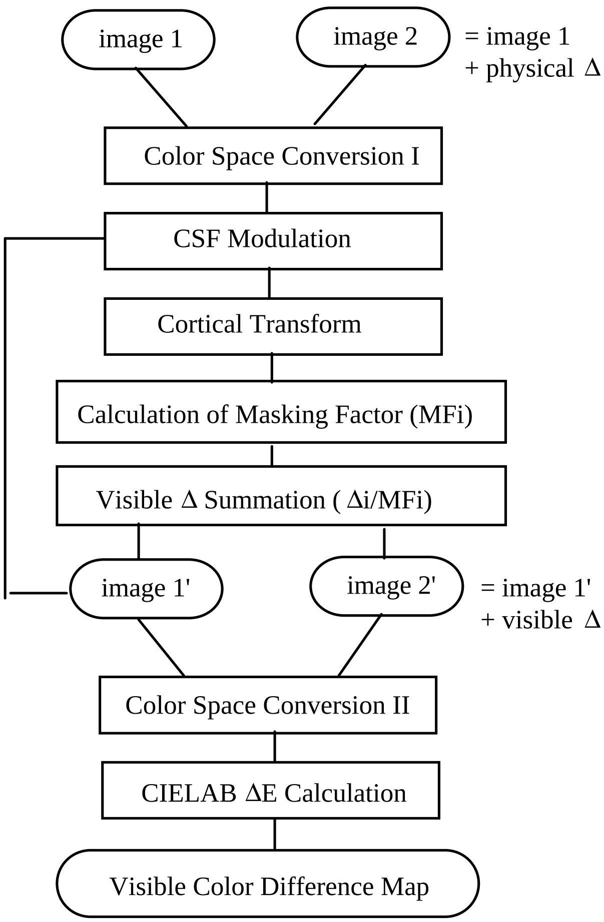

Figure 2: The flowchart of the CVDM.

Test of the Color Difference Model

The main purpose of the test is to examine the masking effect of the CVDM and the compatibility of the CVDM and the S-CIELAB models. The reference image is a chapel image (256 by 256, Figure 3(A)). This image contains both low spatial frequency information (in the sky area) and high spatial frequency information (in the grill area of the chapel). The viewing distance and the spatial resolution of the image are set to be 18 inches and 75 dpi. The Nyquist frequency of the image is thus 11.8 cpd. Four kinds of test images are used to test the CVDM: blurred image, white noise image, grating noise image, and compression/decompression image. Due to the limited space of the paper, only the test using grating image (Figure 3(B)) is shown here. Figure 3(C) shows the prediction of the CVDM. The prediction of the S-CIELAB model is included as a comparison (Figure 3(D)). In the output ΔE maps, a bright area represents a high ΔE value, while a dark part represents a low ΔE value. The scales of the two ΔE maps are the same, allowing direct comparison of the predictions of the two models.

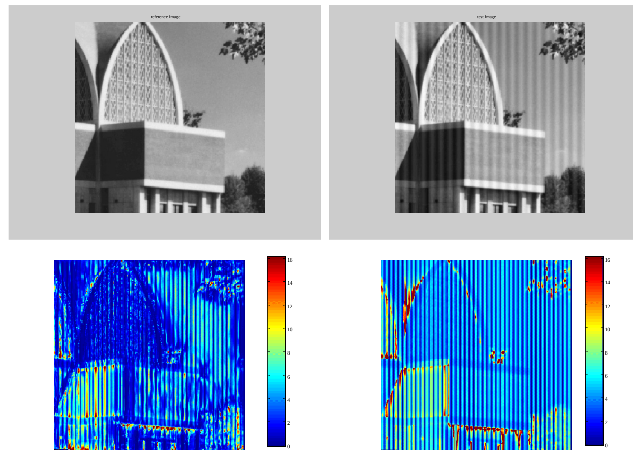

Figure 3: Simulation test of the CVDM with a grating image. (A: top left) Reference image; (B: top right) Grating test image; (C: bottom left) Prediction of the CVDM; (D: bottom right) Prediction of the S-CIELAB model.

The grating test image is formed by adding a sine-wave grating to each of the R, G, & B images of the reference image. A visual examination of the test image shows that the grating pattern in the sky area is more obvious than the grating pattern in the grill area. This is expected because the vertical high-spatial-frequency image content in the grill area masks the grating pattern, and hence reducing the strength of the grating. The prediction of the CVDM is shown in Figure 3(C), which shows a reduced grating modulation in the grill area while a strong grating modulation in the sky area. Figure 3(D) gives the prediction of the S-CIELAB model. It shows the same grating amplitude in both the grill area and the sky area. It is obvious that the prediction of the CVDM is more accurate than that of the S-CIELAB model.

In general, the predictions of the CVDM are more consistent with the visual impression than that of the S-CIELAB model over the four test images. The two models produce very similar results when there is no image content to mask the noise, such as in the sky area. When the image is complicated and contains high-spatial-frequency masking content, such as the grill area, the prediction of the CVDM is more accurate than that of the S-CIELAB model.

Conclusion

The CVDM is developed based on the two previously published visual difference models: the Visible Difference Predictor (VDP) model and the Spatial-CIELAB model (S-CIELAB). The CVDM extents the VDP model in that it can be used to evaluate the visual difference of color images.

The output of the CVDM is compatible with CIE standard CIELAB ΔE, which makes it easy to compare the predictions of the CVDM to that of other color appearance models. The CVDM is superior to the S-CIELAB model in that it incorporates the masking effect and the multiresolution detection mechanisms. Incorporating the two additional factors makes the CVDM work well with complex color images.

References

1. Daly, S., *Digital images and Human Vision*, ed. Watson, A. B., Cambridge: The MIT press, 179 (1993).
2. Zhang, X-M and Wandell, B. A., *Proceedings of the SID Symposiums* (1996).
3. Poirson, A. B. and Wandell, B. A., *J. Opt. Soc. Am. A* , **10**, 2458 (1993).
4. Wandell, B. A., *Proc. Natl. Acad. Sci. USA*, **90**, 9778 (1993).
5. Poirson, A. B. and Wandell, B. A., *Vision Res.*, **36**, 515 (1996).
6. Bauml, K-H & Wandell, B. A., *Vision Res.*, **36**, 2849 (1996).
7. Zhang, X-M, Farrell, J. E. and Wandell, B. A., *Proceedings of the SID Symposiums* (1997).
8. Zhang, X-M, Silverstein, D. A., and Farrell, J. E., *IS&T/SPIE electronic imaging 97* (1997).
9. Wandell, Foundations of Vision, Sunderland: Sinauer Associates, Inc., 247 (1995).
10. Kaiser, P. K. and Boynton, R. M., *Human Color Vision*, 2nd ed., Washington: Optical Society of America (1996).
11. Mullen, K.T., *J. Physiol.* **359**, 381 (1985).

[+] Present address, Visual Sciences Center, University of Chicago, Chicago, IL 60637

[*] Present address, Sharp Labs of America, Inc., Camas, WA 98607