Image appearance modeling

Mark D. Fairchild and Garrett M. Johnson[*]
Munsell Color Science Laboratory, Chester F. Carlson Center for Imaging Science,
Rochester Institute of Technology, Rochester, NY, USA 14623-5604

ABSTRACT

Traditional color appearance modeling has recently matured to the point that available, internationally-recommended models such as CIECAM02 are capable of making a wide range of predictions to within the observer variability in color matching and color scaling of stimuli in somewhat simplified viewing conditions. It is proposed that the next significant advances in the field of color appearance modeling will not come from evolutionary revisions of these models. Instead, a more revolutionary approach will be required to make appearance predictions for more complex stimuli in a wider array of viewing conditions. Such an approach can be considered image appearance modeling since it extends the concepts of color appearance modeling to stimuli and viewing environments that are spatially and temporally at the level of complexity of real natural and man-made scenes. This paper reviews the concepts of image appearance modeling, presents iCAM as one example of such a model, and provides a number of examples of the use of iCAM in still and moving image reproduction.

Keywords: color appearance, image appearance, image quality, vision modeling, image rendering

1. INTRODUCTION

The fundamental theme of this research can be considered image measurement and the application of those measurements to image rendering and image quality evaluation. Consideration of the history of image measurement helps set the context for the formulation and application of image appearance models, a somewhat natural evolution of color appearance, spatial vision, and temporal vision models. Early imaging systems were either not scientifically measured at all, or measured with systems designed to specify the variables of the imaging system itself. For example, densitometers were developed for measuring photographic materials with the intent of specifying the amounts of dye or silver produced in the film. In printing, similar measurements would be made for the printing inks as well as measures of the dot area coverage for halftone systems. In electronic systems like television, system measurements such as signal voltages were used to quantify the imaging system. As imaging systems evolved in complexity and openness, the need for device-independent image measures became clear.

1.1 Image Colorimetry

Electronic imaging systems, specifically the development of color television, prompted the first application of device-independent color measurements of images. Device-independent color measurements are based on the internationally-standardized CIE system of colorimetry first developed in 1931. CIE colorimetry specifies a color stimulus with numbers proportional to the stimulation of the human visual system independent of how the color stimulus was produced. The CIE system was used very successfully in the design and standardization of color television systems (including recent digital television systems).

Application of CIE colorimetry to imaging systems became much more prevalent with the advent of digital imaging systems and, in particular, the use of computer systems to generate and proof content ultimately destined for other media such as print. As color-capable digital imaging systems (from scanners and cameras, through displays, to various hardcopy output technologies) became commercially available in the last two decades, it was quickly recognized that device-dependent color coordinates (such as monitor RGB and printer CMYK) could not be used to specify and reproduce color images with accuracy and precision. An additional factor was the open-systems nature of digital imaging in which the input, display, and output devices might be produced by different manufacturers and one source could not control color through the entire process. The use of CIE colorimetry to specify images across the various devices promised to solve some of the new color reproduction problems created by open, digital systems. The flexibility

[*] mdf@cis.rit.edu, garrett@cis.rit.edu, www.cis.rit.edu/mcsl

of digital systems also made it possible and practical to perform colorimetric transformations on image data in attempts to match the colors across disparate devices and media.

Research on imaging device calibration and characterization has spanned the range from fundamental color measurement techniques to the specification of a variety of devices including CRT, LCD, and projection displays, scanners and digital cameras, and various film recording and print media. Some of the concepts and results of this research have been summarized by Berns.[1] Such capabilities are a fundamental requirement for research and development in color and image appearance. Research on device characterization and calibration provides a means to tackle more fundamental problems in device-independent color imaging. For example, conceptual research on design and implementation of device-independent color imaging,[2] gamut mapping algorithms to deal with the reproduction of desired colors that fall outside the range that can be obtained with a given imaging device,[3] and computer graphics rendering of high-quality spectral images that significantly improve the potential for accurate color in rendered scenes.[4] This type of research built upon, and contributed to, research on the development and testing of color appearance models for cross-media image reproduction.

1.2 Color Appearance

Unfortunately, fundamental CIE colorimetry does not provide a complete solution. CIE colorimetry is only strictly applicable to situations in which the original and reproduction are viewed in identical conditions. By their very nature, the images produced or captured by various digital systems are examined in widely disparate viewing conditions, from the original captured scene, to a computer display in a dim room, to printed media under a variety of light sources. Thus color appearance models were developed to extend CIE colorimetry to the prediction of color appearance (not just color matches) across changes in media and viewing conditions (not just within a single condition). Color appearance modeling research applied to digital imaging systems was very active throughout the 1990s culminating with the recommendation of the CIECAM97s model in 1997[5] and its revision, CIECAM02, in 2002.[6] Details on the evolution, formulation, and application of color appearance models can be found in Fairchild.[7] The development of these models was also enabled by visual experiments performed to test the performance of published color appearance models in realistic image reproduction situations.[8] Such research on color appearance modeling in imaging applications naturally highlighted the areas that are not adequately addressed for spatially complex image appearance and image quality problems.

1.3 Image Appearance and Quality

Color appearance models account for many changes in viewing conditions, but are mainly focused on changes in the color of the illumination (white point), the illumination level (luminance), and surround relative luminance. Such models do not directly incorporate any of the spatial or temporal properties of human vision and the perception of images. They essentially treat each pixel of an image (and each frame of a video) as completely independent stimuli. While color appearance modeling has been successful in facilitating device-independent color imaging and is incorporated into modern color management systems, there remains significant room for improvement. To address these issues with respect to spatial properties of vision and image perception (localized adaptation and spatial filtering) and image quality, the concept of image appearance models has been recently introduced and implemented.[9,10] These models combine attributes of color appearance models with attributes of spatial vision models that have been previously used for image quality metrics in an attempt to further extend the capabilities of color appearance models. Historically color appearance models largely ignored spatial vision (e.g., CIECAM97s) while spatial vision models for image quality largely ignored color.[11,12] The goal in developing an image appearance model was to bring these research areas together to create a single model applicable to image appearance, image rendering, and image quality specifications and evaluations. One such model for still images, referred to as iCAM, has recently been published by Fairchild and Johnson[11] and this paper includes an initial extension to the temporal domain to examine digital video appearance. This model was built upon previous research in uniform color spaces,[13] the importance of image surround,[14] algorithms for image difference and image quality measurement,[15,16] insights into observers eye movements while performing various visual imaging tasks and adaptation to natural scenes,[17,18] and an earlier model of spatial and color vision applied to color appearance problems and high-dynamic-range (HDR) imaging.[19] The structure of the iCAM model, examples of its implementation for image appearance, and its extension to video appearance are presented below.

1.4 Still & Moving Image Appearance and Quality

Visual adaptation to scenes and images is not only spatially localized according to some low-pass characteristics, but also temporally localized in a similar manner. To predict the appearance of digital video sequences, particularly those of high-dynamic range, the temporal properties of light and chromatic adaptation must be considered. To predict the quality (or image differences) of video sequences, temporal filtering to remove imperceptible high-frequency temporal modulations (imperceptible "flicker") must be added to the spatial filtering that removes imperceptible spatial artifacts (*e.g.*, noise or compression artifacts). This paper describes a first attempt at spatial adaptation for video sequences. Future research is planned to enhance this first attempt, to implement spatio-temporal filtering, and to evaluate both psychophysically.

It is easy to illustrate that adaptation has a significant temporal low-pass characteristic. For example, if one suddenly turns on the lights in a darkened room (as upon first awakening in the morning), the increased illumination level is at first dazzling to the visual system, essentially overexposing it. After a short period of time, the visual system adapts to the new, higher level of illumination and normal visual perception becomes possible. The same is true when going from high levels of illumination to low levels (imagine driving into a tunnel in the daytime). Fairchild and Reniff[20] and Rinner and Gegenfurtner[21] have made detailed measurements of the time-course of chromatic adaptation. The Fairchild and Reniff results were used to create a temporal integration function to be applied to the XYZ adaptation image in the iCAM model. Briefly, the adaptation image depends not only on a spatially low-pass version of the current image frame, but a temporally low-pass version of the frames viewed in the previous ten seconds (enough time to capture most of the temporal effect). Thus a bright frame viewed immediately after a series of dark frames will appear (and be rendered) significantly brighter than the same frame viewed after a period of adaptation to similarly illuminated frames. There are two advantages to such processing. One is that the appearance of the rendered video mimics that of human perception and the second is that HDR video or cinema sequences can be rendered on low-dynamic-range displays (such as streaming video to an LCD on a laptop).

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. This research is relevant to the extension of iCAM to measure video differences, but has been undertaken with a very different objective. 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 iCAM work summarized above[10,16] has had a different objective with respect to image quality. Instead of focusing on threshold differences in quality, the focus has been on the prediction of image quality scales (*e.g.*, scales of sharpness, contrast, graininess) for images with changes well above threshold. Such suprathreshold image differences are a different domain of image quality research based on image appearance that separate the iCAM model from previous image quality models.

Likewise, a similar situation exists in the area of video quality metrics. Metrics have been published to examine the probability of detection of artifacts in video (*i.e.*, threshold metrics), but there appears to be no models of video image appearance designed for rendering video and predicting the magnitudes of perceived differences in video sequences. The latter is the goal of the extension of iCAM. 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 the proposed extensions to the iCAM model.

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[22] and more fully disclosed in other publications.[23] It is based on the multi-scale model of spatial vision published by Lubin[12,24] with some extensions for color processing and temporal variation. The Lubin model is similar in nature to the Daly model mentioned above 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 and luminance adaptation as are included in the iCAM model. 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. The Sarnoff model also is not designed for rendering video. This is not a criticism of the model formulation, but an illustration of how the objective of the Sarnoff JND model is significantly different from that of the iCAM 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[25] and Watson *et al.*[26] 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 and therefore also cannot be used for video rendering. Again, this is not a shortcoming, but rather a property of the design objectives for the DVQ model.

In summary, while there is significant literature available on the visual modeling of digital video quality, it remains sparse and the available models were designed with objectives that differ significantly from those of this research. The extensions of the iCAM model for digital video applications will include temporal aspects of image adaptation and appearance in addition to suprathreshold video image difference metrics. Such extensions will enable new types of video rendering, artistic and visually veridical re-rendering of video and cinema content into different media, and prediction of perceived differences in video sequences at suprathreshold levels in both the spatial and temporal domain.

2. THE iCAM FRAMEWORK

Figure 1 presents a flow chart of the general framework for the iCAM image appearance model as applied to still images originally presented by Fairchild and Johnson.[10] A description of the model along with example images and code can be found at <www.cis.rit.edu/mcsl/iCAM>. For input, the model requires colorimetrically characterized data for the image (or scene) and surround in absolute luminance units. The image is specified in terms of relative CIE XYZ tristimulus values. The adapting stimulus is a low-pass filtered version of the CIE XYZ image that is also tagged with absolute luminance information necessary to predict the degree of chromatic adaptation. The absolute luminances (Y) of the image data are also used as a second low-pass image to control various luminance-dependant aspects of the model intended to predict the Hunt effect (increase in perceived colorfulness with luminance) and the Stevens effect (increase in perceived image contrast with luminance). Lastly, a low-pass, luminance (Y) image of significantly greater spatial extent is used to control the prediction of image contrast that is well-established to be a function of the relative luminance of the surrounding conditions (Bartleson and Breneman equations). Refer to Fairchild[7] for a full discussion of the various image appearance effects mentioned above and detailed specifications of the data required. The specific low-pass filters used for the adapting images depend on viewing distance and application. Additionally, in some image rendering circumstances it might be desirable to have different low-pass adapting images for luminance and chromatic information to avoid desaturation of the rendered images due to local chromatic adaptation (decrease in visual sensitivity to the color of the stimulus). This is one example of application dependence. Local chromatic adaptation might be appropriate for image-difference or image-quality measurements, but inappropriate for image-rendering situations.

The first stage of processing in iCAM is to account for chromatic adaptation. The chromatic adaptation embedded in the recently-published CIECAM02 model[6] has been adopted in iCAM since it was well researched and established to have excellent performance with all available visual data. It is also a relatively simple chromatic adaptation model amenable to image-processing applications. The chromatic adaptation model is a linear von Kries normalization of RGB image signals to the RGB adaptation signals derived from the low-pass adaptation image at each pixel location. The RGB signals are computed using a linear transformation from XYZ to RGB derived by CIE TC8-01 in the formulation of CIECAM02. This matrix transformation has come to be called the M_{CAT02} matrix, where CAT stands for chromatic adaptation transform. The von Kries normalization is further modulated with a degree-of-adaptation factor, D, that can vary from 0.0 for no adaptation to 1.0 for complete chromatic adaptation. An equation is provided in the CIECAM02

formulation, and used in iCAM, for computation of D for various viewing conditions. Alternatively the D factor can be established manually. It should be noted that, while the adaptation transformation is identical to that in CIECAM02, the iCAM model is already significantly different since it uses spatially-modulated image data as input rather than single color stimuli and adaptation points. It also differs completely in the remainder of the formulation although using CIECAM02 equations where appropriate. One example of this is the modulation of the absolute-luminance image and surround luminance image using the F_L function from CIECAM02. This function, slowly varying with luminance, has been established to predict a variety of luminance-dependent appearance effects in CIECAM02 and earlier models. Since the function has been established and understood, it was also adopted for the early stages of iCAM. However, the manner in which the F_L factor is used in CIECAM02 and iCAM are quite different.
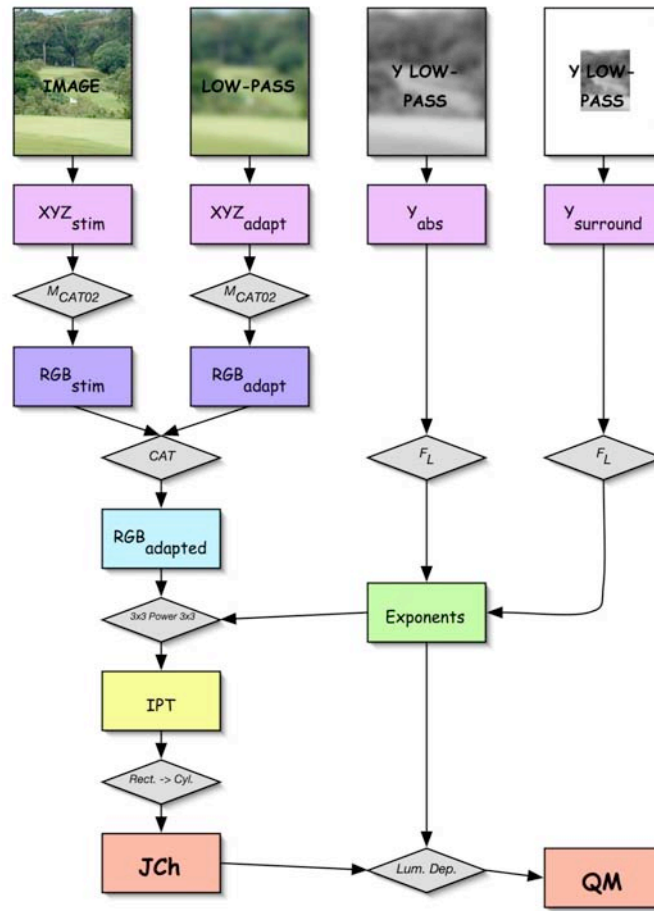

Figure 1. Flow chart of the iCAM image appearance model.

The next stage of the model is to convert from RGB signals (roughly analogous to cone signals in the human visual system) to opponent-color signals (light-dark, red-green, and yellow-blue; analogous to higher-level encoding in the human visual system) that are necessary for constructing a uniform perceptual color space and correlates of various appearance attributes. In choosing this transformation, simplicity, accuracy, and applicability to image processing were the main considerations. The color space chosen was the IPT space previously published by Ebner and Fairchild.[13] The IPT space was derived specifically for image processing applications to have a relatively simple formulation and specifically to have a hue-angle component with good prediction of constant perceived hue (important in gamut-mapping applications). More recent work on perceived hue has validated the applicability of the IPT space. The transformation from RGB to the IPT opponent space is far simpler than the transformations used in CIECAM02. The process involves a linear transformation to a different cone-response space (a different RGB), application of power-function

nonlinearities, and then a final linear transformation to the IPT opponent space (I: light-dark; P: red-green, T: yellow-blue).

The power-function nonlinearities in the IPT transformation are a critical aspect of the iCAM model. First, they are necessary to predict response compression that is prevalent in most human sensory systems. This response compression helps to convert from signals that are linear in physical metrics (*e.g.*, luminance) to signals that are linear in perceptual dimensions (*e.g.*, lightness). The CIECAM02 model uses a hyperbolic nonlinearity for this purpose. The behavior of which is that of a power function over the practical ranges of luminance levels encountered. Secondly, and a key component of iCAM, the exponents are modulated according to the luminance of the image (low-pass filtered) and the surround. This is essentially accomplished by multiplying the base exponent in the IPT formulation by the image-wise computed F_L factors with appropriate normalization. These modulations of the IPT exponents allow the iCAM model to be used for predictions of the Hunt, Stevens, and Bartleson/Breneman effects mentioned above. They also happen to enable the tone mapping of high-dynamic-range images into low-dynamic range display systems in a visually meaningful way (see example in Fig. 4).

For image-difference and image-quality predictions, it is also necessary to apply spatial filtering to the image data to eliminate any image variations at spatial frequencies too high to be perceived. For example, the dots in a printed half-tone image are not visible if the viewing distance is sufficiently large. This computation is dependent on viewing distance and based on filters derived from human contrast sensitivity functions. Since the human contrast-sensitivity functions vary for luminance (band-pass with sensitivity to high frequencies) and chromatic (low pass) information, it is appropriate to apply these filters in an opponent space. Thus in image-quality applications of iCAM, spatial filters are applied in the IPT space. Since it is appropriate to apply spatial filters in a linear-signal space, they are applied in a linear version of IPT prior to conversion into the non-linear version of IPT for appearance predictions. Johnson and Fairchild have recently discussed some of the important considerations for this type of filtering in image-difference applications and specified the filters used based on available visual data.[16] Since the spatial filtering effectively blurs the image data, it is not desirable for image rendering applications in which observers might view the images more closely than the specified viewing distance. The result would be a blurrier image than the original. It is only appropriate to apply these spatial filters when the goal is to compute perceived image differences (and ultimately image quality). This is an important distinction between spatially-localized adaptation (good for rendering and image quality metrics) and spatial filtering (good for image quality metrics, bad for rendering). In image-quality applications, the spatial filtering is typically broken down into multiple channels for various spatial frequencies and orientations. For example, Daly,[11] Lubin,[12] and Pattanaik *et al.*[19] describe such models. More recent results suggest that while such multi-scale and multi-orientation filtering might be critical for some threshold metrics, it is often not necessary for data derived from complex images and for supra-threshold predictions of perceived image differences (one of the main goals of iCAM). Thus, to preserve the simplicity and ease of use of the iCAM model, single-scale spatial filtering with anisotropic filters was adopted.

Once the IPT coordinates are computed for the image data, a simple coordinate transformation from rectangular to cylindrical coordinates is applied to obtain image-wise predictors of lightness (J), chroma (C), and hue angle (h). Differences in these dimensions can be used to compute image difference statistics and those used to derive image quality metrics. In some instances, correlates of the absolute appearance attributes of brightness (Q) and colorfulness (M) are required. These are obtained by scaling the relative attributes of lightness and chroma with the appropriate function of F_L derived from the image-wise luminance map.

For image rendering applications, the main focus of this paper, it is necessary to take the computed appearance correlates (JCh) and then render them to the viewing conditions of a given display. The display viewing conditions set the parameters for the inversion of the IPT model and the chromatic adaptation transform (all for an assumed spatially uniform display adaptation typical of low-dynamic-range output media). This inversion allows the appearance of original scenes or images from disparate viewing conditions to be rendered for the observer viewing a given display. One important application of such rendering is the display of high-dynamic-range (HDR) image data on typical displays.
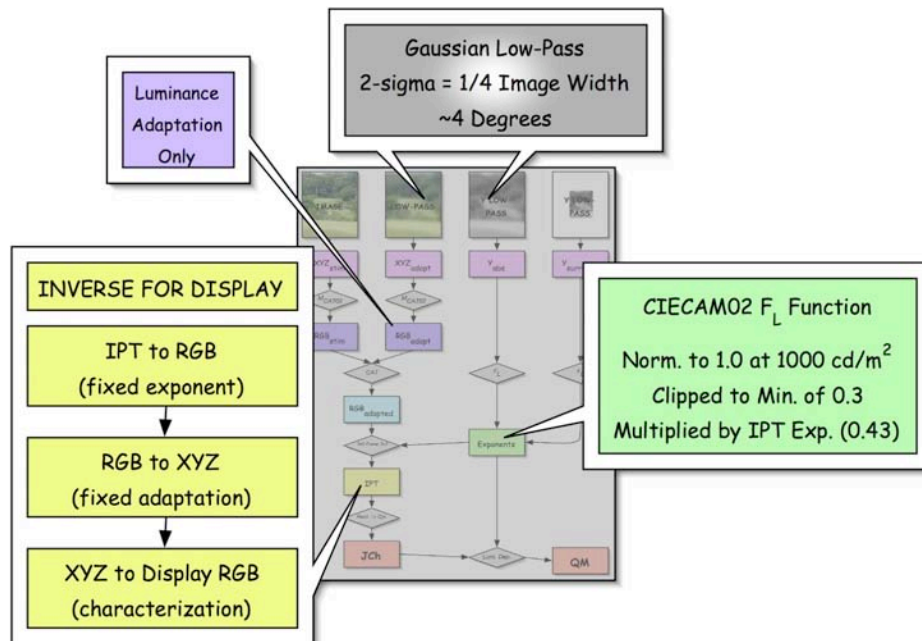
3. IMAGE APPEARANCE APPLICATIONS (RENDERING)

Figure 2. Implementation of iCAM for tone mapping of HDR images.

Figure 2 illustrates the extensions to the basic iCAM model required to complete an image rendering process necessary for HDR image tone mapping. The components essential in this process are the inversion of the IPT model for a single set of spatially constant viewing conditions (the display) and the establishment of spatial filters for the adapting stimuli used for local luminance adaptation and modulation of the IPT exponential nonlinearity. While the derivation of optimal model settings for HDR image rendering is still underway, quite satisfactory results have been obtained using the settings described below.

4. IMAGE QUALITY APPLICATIONS (DIFFERENCE PERCEPTIBILITY)
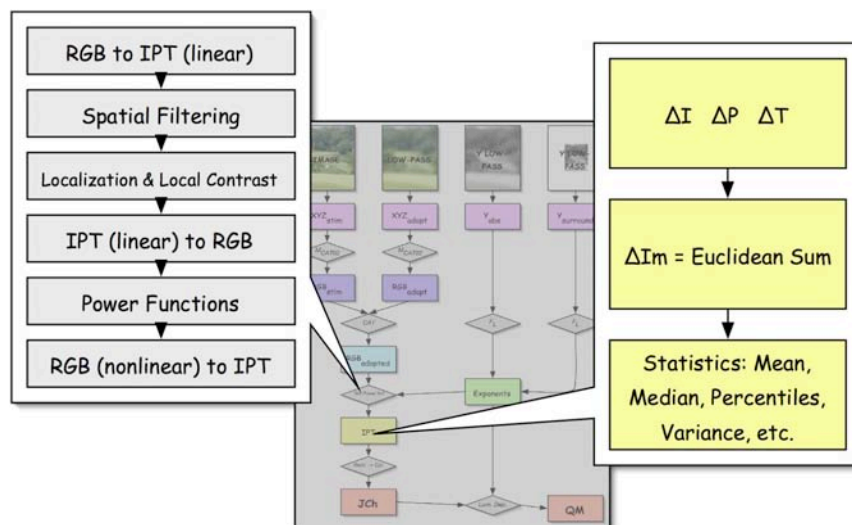

Figure 3. Implementation of iCAM for image difference and image quality metrics.

A slightly different implementation of iCAM is required for image quality applications in order to produce image maps representing the magnitude of perceived differences between a pair of images. In these applications, viewing-distance-dependent spatial filtering is applied in a linear IPT space and then differences are computed in the normal nonlinear IPT space. Euclidean summations of these differences can be used as an overall color difference map and then various summary statistics can be used to predict different attributes of image difference and quality. This process is outlined in Fig. 3 and described more fully in Johnson and Fairchild.[27]

5. IMAGE RENDERING EXAMPLES

The iCAM model has been successfully applied to prediction of a variety of color appearance phenomena such as chromatic adaptation (corresponding colors), color appearance scales, constant hue perceptions, simultaneous contrast, crispening, spreading, and image rendering.[10] One of the most interesting and promising applications of iCAM is to the rendering of high-dynamic-range (HDR) images to low-dynamic-range display systems. HDR image data are quickly becoming more prevalent. Historically HDR images were obtained through computer graphics simulations computed with global-illumination algorithms (*e.g.*, ray tracing or radiosity algorithms) or through the calibration and registration of images obtained through multiple exposures. Real scenes, especially those with visible light sources, often have luminance ranges of up to six orders of magnitude. More recently, industrial digital imaging systems have become commercially available that can more easily capture HDR image data. It is also apparent that consumer digital cameras will soon be capable of capturing greater dynamic ranges. Unfortunately display and use of such data are difficult and will remain so since even the highest-quality displays are generally limited in dynamic range to about two orders of magnitude. One approach is to interactively view the image and select areas of interest to be viewed optimally within the display dynamic range. This is only applicable to computer displays and not appropriate for pictorial imaging and printed output. Another limitation is the need for capability to work with greater than 24-bit (and often floating point) image data. It is desirable to render HDR pictorial images onto a display that can be viewed directly (no interactive manipulation) by the observer and appear similar to what the observer would perceive if the original scene was viewed. For printed images, it is not just desirable, but necessary. Pattanaik *et al.*[19] review several such HDR rendering algorithms and it is worth noting that several papers were presented on the topic at the most recent SIGGRAPH meeting, illustrating continued interest in the topic.

Figure 4. Three HDR images from <www.debevec.org>. The leftmost column illustrates linear rendering of the image data, the middle column illustrates manually-optimized power-function transformations, and the rightmost column represents the automated output of the iCAM model implemented for HDR rendering (see Fig. 2).

Since iCAM includes spatially-localized adaptation and spatially-localized contrast control, it can be applied to the problem of HDR image rendering. This is not surprising since the fundamental problem in HDR rendering is to reproduce the appearance of an HDR image or scene on a low-dynamic-range display. Since the encoding in our visual system is of a rather low dynamic range, this is essentially a replication of the image appearance processing that goes on in the human observer and is being modeled by iCAM. Figure 4 illustrates application of the iCAM model to HDR images obtained from Debevec <www.debevec.org>. The images in the left column of Fig. 4 are linear renderings of the original HDR data normalized to the maximum presented simply to illustrate how the range of the original data exceeds a typical 24-bit (8-bits per RGB channel) image display. For example, the memorial image data (top row) have a dynamic range covering about six orders of magnitude since the sun was behind one of the stained-glass windows. The middle column of images represents a typical image-processing solution to rendering the data. One might consider a logarithmic transformation of the data, but that would do little to change the rendering in the first column. Instead the middle column was generated interactively by finding the optimum power-function transformation (also sometimes referred to as gamma correction; note that the linear images in the first column are already gamma corrected). For these images, transformations with exponents, or gammas, of approximately 1/6 (as opposed to 1/1.8 to 1/2.2 for typical displays) were required to make the image data in the shadow areas visible. While these power-function transformations do make more of the image-data visible, they required user interaction, tend to wash out the images in a way not consistent with the visual impression of the scenes, and introduce potentially-severe quantization artifacts in shadow regions. The rightmost column of images shows the output of the iCAM model with spatially-localized adaptation and contrast control (as shown in Fig. 2). These images both render the dynamic range of the scene to make shadow areas visible and retain the colorfulness of the scene. The resulting iCAM images are quite acceptable as reproductions of the HDR scenes (equivalent to the result of dodging and burning historically done in photographic printing). It is also noteworthy that the iCAM-rendered images were all computed with an automated algorithm mimicking human perception with no user interaction.
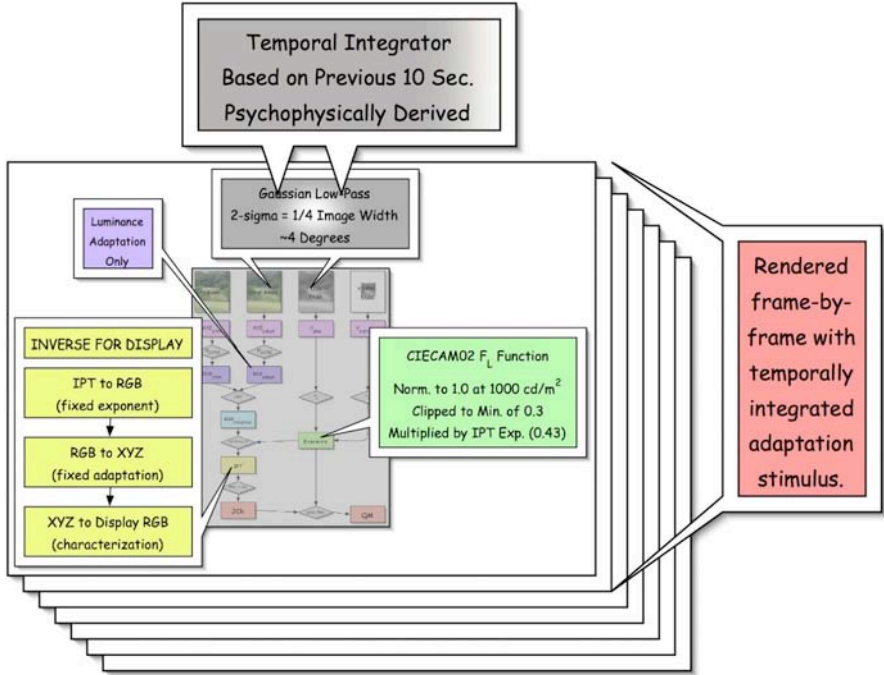
6. DIGTAL VIDEO RENDERING

Figure 5. Implementation of iCAM for tone mapping of HDR video sequences. The temporal integrator is given in Eq. 1.

The extension of iCAM to digital video applications requires implementation of a temporally low-pass function to model the time-course of chromatic and light adaptation for rendering applications and the extension of the spatial filters to

spatio-temporal filters for image difference and quality applications. Only video rendering, and thus the temporal properties of adaptation, are addressed in this paper. Fairchild and Reniff[20] collected data on the time-course of chromatic adaptation to image displays and found that it was essentially complete after about 2 minutes with much of process complete in a few seconds. Further analysis of their data suggested that adequate video rendering could be accomplished by computing the adaptation for each video frame based on the previous 10 sec. of video. To derive a temporal integration function, the degree-of-adaptation data as a function of time after a sharp transition of the adapting stimulus of Fairchild and Reniff[20] was examined. The visual data were described using a sum-of-two-exponentials function. An average function was derived for all of the viewing conditions used in the experiments, flipping the function into the negative-time domain to examine the effect of previous exposures, and then taking the derivative of the function (since the collected data were effectively a cumulative integral. Examination of this function shows that the value at negative 10 sec. is 0.75% of the value at 0 sec. and thus the decision was made that 10 sec. of integration was satisfactory for practical applications. Equation 1 is the final temporal integration function, $AW(f)$ for adapting weight, expressed in terms of numbers of frames with an assumption of 30 frames per second ($f = 0$ for the current video frame and $f = -300$ for the frame that passed 10 sec. ago) and normalized to unit area. The implementation of this temporal integrator is illustrated in Fig. 5.

$$AW(f) = 0.000261e^{0.000767f} + 0.0276e^{0.0297f} \qquad (1)$$
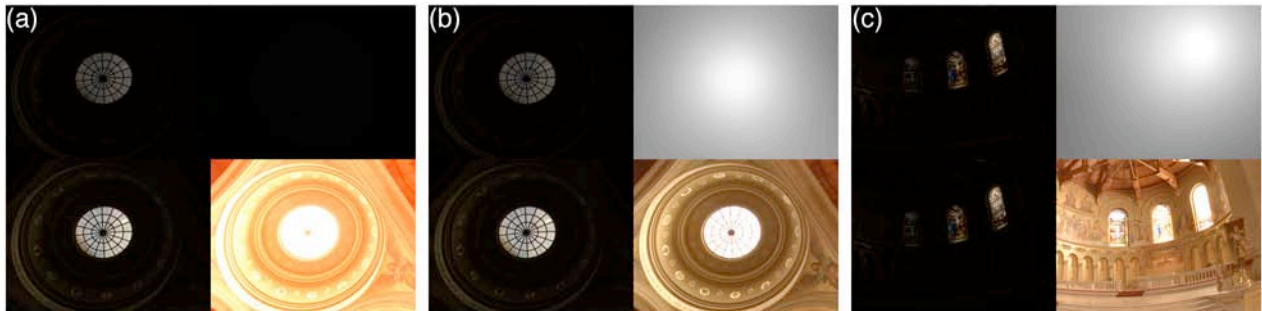

Figure 6. Frames from a video sequence rendered with iCAM extended as shown in Fig. 5. See text for full explanation. (a) First frame of image data after dark adaptation, (b) 10 sec. after the initial exposure, and (c) final frame of the sequence. Upper right sub-frames show the spatially and temporally integrated adapting luminance image and the lower right sub-frames show the iCAM rendered video frames.

A simple example HDR video sequence was created by scanning a small frame through the Debevec's HDR memorial scene. The sequence begins with 10 sec. of darkness to set the model to dark adaptation. There is then an abrupt transition to a view of the round window at the top of the memorial scene. This view is fixated for 10 sec. to illustrate the temporal changes in adaptation (applied both to the local luminance adaptation and local contrast adaptation mechanisms). The sequence then scans through the scene to show other transitions in appearance. Figure 6 shows 3 frames extracted from the video sequence. Each frame is actually a composite of 4 sub-frames. The upper left sub-frame is the linearly-rendered HDR image data assuming no frame-by-frame gain control. The lower left sub-frame is also linearly-rendered, but includes a frame-by-frame gain control. This sub-frame illustrates that even small segments of the original scene often contain HDR image data. The upper right sub-frame shows the luminance channel of the temporally-integrated (and spatially low pass) adaptation image. This is the image used to set the luminance adaptation and the IPT exponents. Lastly, the lower right sub-frame shows the fully rendered video processed through the spatial and temporal iCAM model. The three frames show (a) the first frame of image data immediately following transition from the dark frames, (b) 10 sec. later after adaptation to the same view has stabilized, and (c) the final frame of the sequence showing a typical adaptation state during a scan through the scene. Note that the rendered sub-frame in (a) is extremely bright as is typically witnessed upon entering a brightly illuminated scene (or upon opening one's eye's after a period of dark adaptation) while the adaptation stimulus is dark since there is no prior exposure. Frame (b) shows how the adaptation stimulus has built up over the previous 10 sec. and it's effect as witnessed in the rendered sub-frame, which is similar to a steady-state view of the still image as given in Fig. 4.

7. CONCLUSIONS

Advances in imaging and computing technologies along with increased knowledge of the function and performance of the human visual system have allowed for the integration of models of color, spatial, and temporal vision to create a new type of color appearance model, referred to as an image appearance model. Such models show promise in a variety of applications ranging from image difference and image quality metrics to the rendering of image data. This paper described the framework of one example of an image appearance model referred to as iCAM and illustrated its applicability to HDR image tone mapping along with initial efforts to extend the model to video appearance and quality applications. Future efforts will be directed at completion of the spatio-temporal filters required for video difference metrics, the collection of more psychophysical data on image and video appearance and differences, and the formulation of specific iCAM algorithms for various applications.

8. REFERENCES

1. R.S. Berns, A generic approach to color modeling, *Color Research and Application* **22,** 318-325 (1997).
2. M.D. Fairchild, Some hidden requirements for device-independent color imaging, *SID International Symposium,* San Jose 865-868 (1994).
3. G.J. Braun and M.D. Fairchild, General-purpose gamut-mapping algorithms: Evaluation of contrast-preserving rescaling functions for color gamut mapping, *Journal of Imaging Science and Technology* **44**, 343-350 (2000).
4. G.M. Johnson and M.D. Fairchild, Full-spectral color calculations in realistic image synthesis, *IEEE Computer Graphics & Applications* **19:4**, 47-53 (1999).
5. CIE, The CIE 1997 Interim Colour Appearance Model (Simple Version), CIECAM97s, *CIE Pub. 131* (1998).
6. N. Moroney, M.D. Fairchild, R.W.G. Hunt, C.J Li, M.R. Luo, and T. Newman, The CIECAM02 color appearance model, *IS&T/SID 10th Color Imaging Conference*, Scottsdale, 23-27 (2002).
7. M.D. Fairchild, *Color Appearance Models,* Addison-Wesley, Reading, Mass., (1998).
8. K.M. Braun and M.D. Fairchild, Testing five color appearance models for changes in viewing conditions, *Color Research and Application* **22,** 165-174 (1997).
9. M.D. Fairchild, Image quality measurement and modeling for digital photography, *International Congress on Imaging Science '02*, Tokyo, 318-319 (2002).
10. M.D. Fairchild and G.M. Johnson, Meet iCAM: A next-generation color appearance model, *IS&T/SID 10th Color Imaging Conference*, Scottsdale, 33-38 (2002).
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. 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).
13. F. Ebner, and M.D. Fairchild, Development and testing of a color space (IPT) with improved hue uniformity, *IS&T/SID 6th Color Imaging Conference,* Scottsdale, 8-13 (1998).
14. M.D. Fairchild, Considering the surround in device-independent color imaging, *Color Research and Application* **20,** 352-363 (1995).
15. M.D. Fairchild, Modeling color appearance, spatial vision, and image quality, *Color Image Science: Exploiting Digital Media,* Wiley, New York, 357-370 (2002b).
16. G.M. Johnson and M.D. Fairchild, A top down description of S-CIELAB and CIEDE2000, *Color Research and Application*, in press (2003b).
17. J.S. Babcock, J.B. Pelz and M.D. Fairchild, Eye tracking observers during color image evaluation tasks, *SPIE/IS&T Electronic Imaging Conference*, Santa Clara, in press (2003).
18. M.A. Webster and J.D. Mollon, Adaptation and the color statistics of natural images, *Vision Res.* **37**, 3283-3298 (1997).
19. S.N. Pattanaik, J.A. Ferwerda, M.D. Fairchild, and D.P. Greenberg, A multiscale model of adaptation and spatial vision for image display, *Proceedings of SIGGRAPH 98,* 287-298 (1998).
20. M.D. Fairchild and L. Reniff, Time-course of chromatic adaptation for color-appearance judgements, *Journal of the Optical Society of America A* **12,** 824-833 (1995).
21. O. Rinner and K.R. Gegenfurtner, Time course of chromatic adaptation for color appearance discrimination, *Vision Res.* **40**, 1813-1826 (2000).

22. Sarnoff Corporation, JND: A human vision system model for objective picture quality measurements, Sarnoff Technical Report from www.jndmetrix.com, (2001).
23. ATIS, Objective perceptual video quality measurement using a JND-based full reference technique, *Alliance for Telecommunications Industry Solutions Technical Report T1.TR.PP.75-2001*, (2001).
24. 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).
25. A.B. Watson, Toward a perceptual video quality metric, *Human Vision and Electronic Imaging III*, SPIE Vol. **3299**, 139-147 (1998).
26. A.B. Watson, J. Hu, and J.F. McGowan, DVQ: A digital video quality metric based on human vision, *J. Electronic Imaging* **10**, 20-29 (2001).
27. G.M. Johnson and M.D. Fairchild, Measuring images: Differences, Quality, and Appearance, *SPIE/IS&T Electronic Imaging Conference*, Santa Clara, in press (2003).