

Investigation of the Academy's Image Interchange Framework at RIT

Christopher J. Clark

Rochester Institute of Technology — May 21st, 2010

Abstract. Students of the Rochester Institute of Technology's School of Film and Animation produce hundreds of films each year. The focus of their studies is mainly on its story, direction, and editing. Managing what exactly happens to their scene from when they set up the lights on day one to the day the film screens in the Carlson Auditorium is the focus of this paper. There is often an assumption of "what you see is what you get" when it comes to student filmmaking. However, there is an enormous amount of image processing that occurs within the camera on set, the equipment used during post-production, and the projector at final exhibition. Often this produces unwanted or unexpected results. The solution to this problem is to develop a color management framework as a basis on which students can properly manage the "final look" and maintain the creative intent of their projects.

When film is exposed, processed, printed, and projected, the reproduction of real world color and contrast is not perfectly accurate. It is, however, how people seem to *prefer* to see the world. Across the world, people will most often prefer the "film look," one that has been tweaked for decades by Kodak and other companies.¹ The question becomes whether people enjoy this look for habitual reasons, or if it is truly the most pleasing to the eye. Because film is a subtractive system consisting of cyan, magenta, and yellow dyes, achieving that look with a red-green-blue additive system such as digital capture is a challenge. Digital camera manufacturers have been struggled with this issue for years. The organic nature of film does, however, carry its own limitations on color rendering. This begins a philosophical debate of whether future digital systems should attempt to imitate film or discover a look more pleasing than that.

With the shift to digital capture upon us, the Academy of Motion Picture Arts and Sciences has formed a committee to discuss this question, and develop what is known as the *Image Interchange Framework*. The IIF will be a standard encoding and rendering for digital motion picture mastering. Just as the film imaging chain renders a certain "look" when you simply expose and print a piece of film, this new framework is a digital imaging chain, on top of which directors and cinematographers can place their own artistic intentions. For the past four years, work has been done to develop this standard, but the question of whether to imitate film or go beyond it remains in debate. The current stance is to start with film as an aim, but take advantage of digital imaging capabilities to improve it.

During the course of this research project, a key component of the Image Interchange Framework is developed using the equipment available to a Digital Cinema student at the Rochester Institute of Technology. This component is referred to as an *Input Device Transform (IDT)*, a transform which fits a specific input device into the IIF. Extensive characterization of the Panasonic HVX-200 camera is done in order to build IDTs for scenarios typical of a SOFA student using this camera. Future implementation of a complete framework would allow students to be sure that "what they see is what they'll *really* get," and this project serves as a proof-of-concept to help advance the Academy's work.

1 INTRODUCTION

1.1 The Academy's Image Interchange Framework

1.1.1 Goals

As films began going into digital intermediate for post-production during the 1990s, a decision on the method of encoding digital files from scanned film was necessary. Since the vast majority of theatres were still projecting film, Kodak designed their first digital intermediate system, called Cineon, to encode images based on *reference printing density*. After mastering in this filmic context, a film recorder was responsible for translating reference printing density correctly onto a piece of film for mass duplication. To contain these encoded printing densities in data files, the Digital Picture Exchange (.DPX) format was defined by SMPTE and widely adopted to allow for image interchange between D.I. facilities.

These have been the standards since the introduction of the digital intermediate (D.I.) workflow in the 1990s. Obviously, they have worked. However, the components of these standards are often poorly understood. Hardware and software design for D.I. equipment should be designed with a fundamental understanding of colorimetry, the human visual system, and color appearance modeling. Unfortunately, this is not always the case. However, Digital Cinema Initiatives (DCI) has since provided an intelligent and relieving standard for digital projection and theatre distribution that has a solid base for the Academy's framework to connect with.

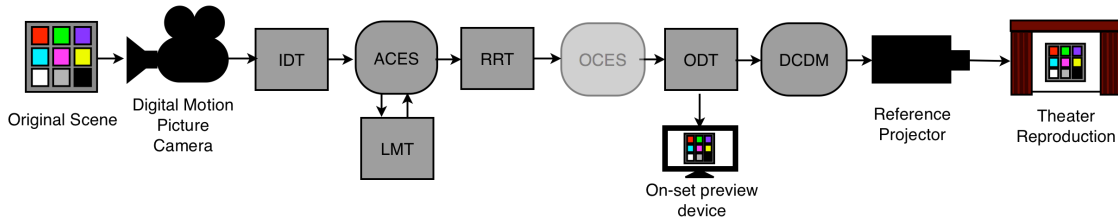
The standards mentioned become increasingly important as footage from digital cameras and CGI are ingested for use in conjunction with scanned film. However, the color differences between different input mediums still exist. While film is relatively consistent between cameras, the controls and automated balancing internal to digital cameras can cause extensive differences and create a headache of metadata in post-production. This creates a problem when material from multiple sources needs to be intercut or composited. Rising costs for time in post-production and loss of image quality is an unfortunate consequence. In 2004, the Academy of Motion Picture Arts and Sciences decided that an industry standard must be established for the digital motion picture mastering. The Advanced Technology Program Subcommittee of the Academy's Science and Technology Council initiated the *Image Interchange Framework* (IIF) project.²

The goal of the IIF project is to establish a robust and unified architecture for digital image and color interchange across the entire motion picture industry. Colors will be encoded according to a clearly defined specification to allow for seamless compositing, and images will be in a common format for interchange. Since 2004, the Academy has defined the Academy Color Encoding Specification which is capable of encoding any color in an unambiguous manner.³ This project investigates the possibility of fitting the equipment of an academic institution into the Academy's framework. The paper provides the background to understand the framework, and the results provide feedback for the Academy as they work towards its completion.

1.1.2 Components

The framework begins with the development of an *Input Device Transform* (IDT), a component that will eventually be integrated into the camera. The IDT accounts for the color and tonescale differences between cameras and attempts to bring them to a common format.⁴ This meeting point is the *Academy Color Encoding Specification* (ACES). A color encoded in ACES is clearly defined and unambiguous, so that the input source does not need to be known in order to work with it. All

Fig. 1: An outline of the AMPAS Image Interchange Framework.



cameras with a proper IDT will produce ACES values that can be interchanged between facilities. Intercutting and compositing material together can be possible without the need for complex transforms. However, the ACES format does not automatically allow for this to be done seamlessly.

It is important to note that ACES *will* preserve the aesthetic characteristics of film and other cameras. For example, the *universal film unbuild* will still unbuild a highly saturated film into highly saturated ACES images. If drastic differences exist and the proper IDT is not used in anticipation of this, trouble can still occur when intercutting material from multiple sources.

Once in ACES, the image is not ready for viewing. Images in ACES space are in “scene state.” This means it is based on the linear amount of light in the scene, and contains the dynamic range of light that the camera or film is capable of capturing, and cannot necessarily be produced by a display device. A scene-based image will not be pleasing to look at in that state, since there have been no tonescale or color adjustments made to deal with changes in viewing conditions. The *Reference Rendering Transform (RRT)* gets the image into a “rendered state.” However, the image in this state must not be limited to any real display device’s capabilities for dynamic range and color. The “rendered state” image is therefore run through an *Output Device Transform (ODT)*.⁵ At this stage, the image is ready for viewing and should be a pleasing reproduction on the display device.

1.2 SoFA Workflows

The School of Film of Animation at RIT (SOFA) has only a handful of possible workflows for students to follow. Depending on their year level, students have access to film cameras, standard-definition video cameras, and high-definition video cameras. For reasons of cost and convenience, video is most often the format of choice. Video cameras are easier to use, cheaper to shoot with, and require less time to process and view footage. Therefore, the high-definition video camera is the typical choice for students of SOFA. Specifically, the SOFA Film/Video cage most often lends out the Panasonic HVX-200 (hereon referred to as “HVX”), which is a professional-consumer grade high-definition video camera. With a 3-CCD capture system capable of HD resolutions up to 1080i, the HVX camera produces crisp, high-definition video, at an affordable price for student filmmaking.⁶ The camera is shown in Figure 2.

Unfortunately, this camera is essentially a “blackbox” of image processing. When a scene is captured by this high-definition video camera, the light is collected by three separate electronic sensors. Each sensor is sensitive to certain areas of the visible wavelength spectrum of light. The three sensors captured ranges of red, green, and blue parts of the light spectrum. The decisions

Fig. 2: Panasonic HVX-200. Seven units are available to students of SoFA.



to have certain shapes and ranges of sensitivity curves is of great importance in camera design. A “colorimetric” camera, capable of capturing colors and reproducing them exactly, is not easy. Luckily, it is also undesired. There are psychological, psychophysical, and physiological processes that affect our preferences for color and tonescale, whether we are conscious of them or not. The overall range of light levels that these sensors can capture, or “exposure latitude,” is tone-mapped in some non-linear, aesthetic fashion and converted to digital code values. These concepts will be explained in further detail in the Background section. This is not information that camera or film manufacturers usually publish about their products, as these spectral sensitivities are a major part of a product’s unique “look.”⁷

The workflow that follows shooting with the HVX is fairly rigid. Students transfer their footage to Apple Final Cut Pro workstations in the Post-Production Labs for editing and color correction. When complete, the project is exported as a Quicktime movie for SOFA screenings. Upon screening, the student has seen their film digitally projected in the Carlson Auditorium. From the student, more often than not, you will hear, “That didn’t look right!” They are correct: it didn’t match the final “look” they had settled on during post-production. The cause of this problem is a lack of proper color management or even simple calibrations of equipment within the workflow.

This applies not only to the HVX workflow described, but to all workflows available to SOFA students. If a student wants to composite his footage with a CGI character that was designed in the 3D lab with an uncalibrated monitor and a different file format, it will be nearly impossible for the image to appear seamless. A unified image interchange framework for all SoFA workflows would solve this problem. While IIF implementation is beyond the scope of this project, this paper will describe the needs and considerations necessarily taken when preparing for this type of implementation.

1.3 Proof of Concept

Presented to the Academy in 2006, the digital color management proposal for the motion picture industry is a comprehensive and unified architecture. As daunting as it may sound, it meets the requirements of every part of the industry. Any type of input medium can fit into its Academy Color Encoding Specification, and any type of output medium can be mastered for distribution. This Image Interchange Framework is designed to be future-proof, with a color encoding specification that supports all possible colors, and compatibility for future input and output devices to be used.

Not only will these new devices be integrated seamlessly, but the system will be able to take full advantage of the new technology’s capabilities.

For example, imagine that a new projector is developed that can achieve 5,000,000:1 contrast ratio and a color gamut that can produce all real-world colors. Since the color encoding specification is based on technology that is not based on any real device, the contrast ratio and gamut of the new device would fit into the system with no loss of quality.

The concept is that any real device that can be characterized can fit into the Academy framework. This project is a “proof-of-concept,” where the possibility of a student workflow at SoFA fitting into the Academy framework is explored. As much as this research will help RIT and universities alike, it should provide feedback to aid in the Academy’s continued development of the framework.

2 BACKGROUND

2.1 Fundamentals of Color

What is *color*? Most people when asked that question would start talking about red, blue, yellow—but *why* are those colors? Another person, might state that certain colors reflect from objects in different ratios. In the most basic scientific explanation, color is the visual sensation that occurs in our brain when different wavelengths of light reflecting or shining into our eyes. Without light, there is no color. From long to short wavelength, the basic color names would be: red, orange, yellow, green, blue, indigo, violet. ROYGBIV is a common way to remember the colors of a rainbow. Looking at a rainbow is looking at white light from the sun split into a spectrum of wavelengths by the water in the atmosphere. As elementary as the word can sound, color has a complex scientific basis.

2.1.1 Colorimetry

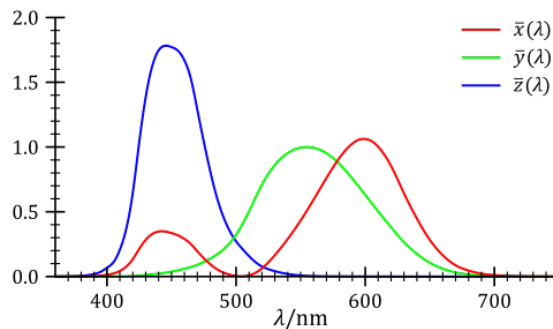
In 1931, an organization called the Commission International de l’Eclairage (CIE) established a way of measuring color. Since Newton’s time, color’s relation to wavelength of light was understood, but there was no way to express colors or color differences with numbers. The CIE conducted a “color-matching experiment” where observers were shown a specific wavelength of light. A separate color directly beside this was under the observer’s control. They had three knobs which controlled three primaries, in this case red, green, and blue. The objective is to use the knobs to make the two colors match. Using the amounts of primaries that each person chose for each wavelength, they constructed *color matching functions*: \bar{r} , \bar{g} , and \bar{b} . Every object has a certain reflectance at each wavelength. When light hits that object, the spectral power distribution of the light source is cascaded with the object’s spectral reflectance to produce what is known as a *color stimulus*. Using this spectral radiance, the \bar{r} , \bar{g} , and \bar{b} functions are cascaded individually and the integral summation of each function yields three *tristimulus values*: RGB. The formulas are shown below, where $\Phi(\lambda)$ is the spectral power of the stimulus.⁸

$$R = \int_{\lambda} \Phi(\lambda) \bar{r}(\lambda) d(\lambda) \quad G = \int_{\lambda} \Phi(\lambda) \bar{g}(\lambda) d(\lambda) \quad B = \int_{\lambda} \Phi(\lambda) \bar{b}(\lambda) d(\lambda)$$

These \bar{r} , \bar{g} , and \bar{b} color-matching functions can be derived through experimentation with any set of primaries. Since the set of primaries from one experiment can be matched using the primaries from

the other system, a linear transform is all that is needed to convert tristimulus values from one set of primaries to the other. Therefore, any three wavelengths of light, or even broadband spectrums can be chosen. At a CIE meeting in 1931, Guild and Wright agreed on this mathematical relationship, but they each had their own data. They decided to make a common set of color-matching functions under certain guidelines. The most pertinent to imaging was that they forced the functions to be all-positive. This was done by choosing a set of primaries that allowed for the description of any physically realizable color, even though the primaries themselves are physically unrealizable. This makes the mathematics behind color and imaging science more straightforward. The other guideline was that they choose primaries so that the \bar{y} function is the CIE 1924 photopic luminous efficiency curve, or $V(\lambda)$. The \bar{y} function now represents our visual sense to luminance information.⁹ These color-matching functions (CMFs) are shown in Figure 3.

Fig. 3: The 1931 CIE 2° standard observer color matching functions.¹⁰ The 2° refers to a 2° viewing angle during the color-matching experiment. The alternative is a set of 10° color-matching functions. Since a viewer is typically looking around an image with complex stimuli, instead of a solid color, the 2° CMFs are the most widely used in the industry.

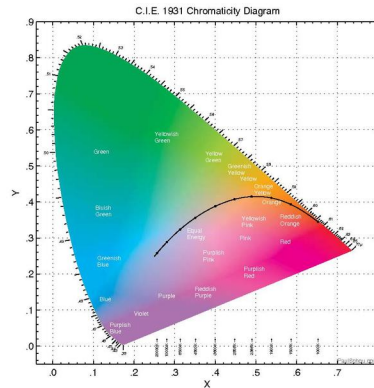


For this reason, these color-matching functions became known as the 1931 2° CIE Standard Observer color-matching functions. The RGB tristimulus formulas simply became the XYZ formulas. The three tristimulus numbers represent the amount of each primary of the color-matching function needed to recreate that color stimulus. Tristimulus values can also be expressed as just two numbers: chromaticity⁸ (shown below). It is a formula which normalizes the tristimulus triplet to eliminate luminance information. By essentially projecting the 3D XYZ space onto a 2D xy space, the chromaticity diagram was invented. Since z can be determined by x and y ($z = 1 - x - y$), the xy chromaticity is a clear definition of tristimulus values. Under identical viewing conditions, colors with the same XYZ tristimulus values will appear to a standard observer with normal color vision to be a visual match.

$$x = \frac{X}{X + Y + Z} \quad y = \frac{Y}{X + Y + Z} \quad z = \frac{Z}{X + Y + Z}$$

Even colors with very different looking spectral reflectances might measure to be the same XYZ values and appear to match. This is known as a *metameric* match. Metamerism occurs when colors with different spectral characteristics appear to match. Illuminant metamerism occurs when two colors do not match under one illuminant, but a different illuminant causes them to match. This occurs because the spectral power distributions of the two light sources are different.

Fig. 4: The chromaticity diagram.¹⁰ While this is indeed the entire spectrum on a 2D space, it can be misleading to think of it that way. Trichromatic color is inherently a 3D space, but placing it on a diagram makes it possible to encompass certain areas, or gamuts, that a device is capable of capturing or displaying.



When these are cascaded with the reflectances of the colors and the CIE color-matching functions, it is possible for the XYZ tristimulus values to match under one illuminant but not the other. Observer metamerism is caused by differences in the cone responses between different observers. For example, two colors might match for one observer but not another. These types of metamerism are unavoidable and often unfortunate, but usually small. However, it is metamerism that *enables* almost all color-imaging equipment to operate. If an apple is to be reproduced on paper, it is unlikely to be able to recreate the exact spectral reflectance of the apple. A CMYK printer has only four dyes. However, with a certain combination of those dyes, it is possible for the same XYZ values to be measured off the paper as those from the apple itself.⁵ This would be a metameric match and a completely accurate reproduction of color for a standard observer.

2.1.2 Color Appearance

There is a distinct difference between *colorimetry* and *color appearance*. Color appearance refers to the final color perception that we make when looking at a color stimulus in a given situation. Consider someone wearing a white shirt on a sunny afternoon. If they were to step inside, one would not see the shirt now as gray. Through what is known as “color constancy,” we adapt to our environment in order to be able to make sense of the visual world through ever-changing conditions.

Distinguishing between *brightness* vs. *lightness* and *colorfulness* vs. *chroma* is important in understanding the effect of color appearance on image reproduction. These are perceptual sensations, but they are each fundamentally different from one another. Let’s imagine it is sunny day and you take a photo of a bright red car. If you looked at the photo inside a dimly lit room, it is unlikely that you would perceive it as *bright* or *colorful*. Lightness and chroma stay constant. This is because lightness and chroma are relative to the level of illuminance that a white object would have in the same environment. Brightness is related to the overall level of luminance hitting an object. The white shirt outside has a higher brightness than it does inside, and we are aware of this perceptually. However, that shirt has a constant lightness based on the measured colorimetry of the shirt, normalized to the overall level of illumination. If the shirt were red, the bright sunlight

would cause the shirt to appear much more colorful than if it were inside. The chroma, however, remains constant because the physical color of the shirt has not changed.

These perceptual factors are not described by a set of XYZ tristimulus values. Colorimetry is simply a numerical description of the color stimulus itself. Under *identical* viewing conditions, tristimulus values will predict whether two stimuli will visually match to a standard observer. CIE colorimetry was developed specifically for this purpose. However, a slight change in the viewing environment will alter the appearance of a color stimulus, even if the colorimetry remains the same.

Thinking about the human visual system as an imaging system helps to put this in perspective. There are typically three stages of a complete imaging system: image capture, signal processing, and image display. A digital camera is an example of a compact imaging system that completes all three stages internally.⁵

1. Lens directs light at the CCD which converts the amount of light into electronic signals.
2. Signal processing to convert these signals to digital code values and demosaick to produce a color image.
3. Code values are sent to the LCD for display.

The human visual system is also a complete imaging system:

1. Light enters the eye, the lens directs the light onto our retina, and the rods and cones send neural signals to the brain.
2. Signal processing occurs through interactions between individual and groups of rods and cones, color memory and experience, situational knowledge.
3. A final “color appearance” is rendered in our brains.

In this sense, colorimetry is only describing the first stage of the human visual system. It describes the color stimulus as our eyes see it. After it hits the retina, there are complex psychophysical and psychological factors that affect a color’s appearance. These effects are numerous and are studied by color scientists for the purpose of creating “color appearance models.” These computational models take, as variables, the various psychophysical and psychological factors to attempt to predict the *final* color that will be sensed by an observer.⁵ While there are textbooks dedicated to the complexity that color appearance models entail, this paper will go into the biggest factors, and really the only factors that affect our perceptions of overall image quality. The most significant part of how an image is perceived by an observer is the viewing environment and state of adaptation that the observer is in.

2.1.3 Observer Adaptation

There are six major viewing condition factors that affect image quality due to the psychophysical adaptations that cause perceptual changes of stimuli.⁵

Viewing flare contribution to stimuli: The amount of viewing flare when viewing an image is a measure of the stray light in the environment that is falling upon the image itself. It is usually expressed as the amount of flare relative to the amount of light from a stimulus that would be perceived as white in the viewing conditions. In addition to the optical flare that washes out your view slightly, this flare light adds lightness to the screen blacks especially, and therefore lowers our overall perception of image contrast and colorfulness.

Stimuli absolute luminance-level adaptation: A color under daylight will appear much brighter to an observer than if the object were taken inside. This is because of the absolute luminance of the object. A reproduction of an outdoor scene will have significantly less luminance than if the observer were standing outside looking at the same scene. Therefore, the observer's perception of contrast and colorfulness goes down when viewing stimuli with significantly lower levels of absolute luminance

Observer chromatic adaptation: The perception of a color stimulus is affected by the observer's state of chromatic adaptation. An observer sitting indoors with tungsten lighting will adapt and be able to tell that a piece of paper is white, even though the colorimetry of that white piece of paper is significantly more yellow than if it were measured outside under daylight, where it would measure more blue. An observer "adaptive white" refers to the chromaticity of a color stimuli that the observer would judge to be perfectly achromatic. This will affect the observer's sense of neutrals and affect the overall color appearance of an image.

Observer lateral-brightness adaptation: Receptive fields on our retinas contain opponency mechanisms, which means that signals are positive and negative. Lateral-brightness adaptation occurs when the gain factors of these opponent signals change depending on the light hitting the surrounding receptive fields. Because of this, an image will have less apparent contrast when the area surrounding the image are dark. If the surround is lighter, the image will appear to have a higher contrast.

Observer general-brightness adaptation: This adaptation occurs based on the overall level of illumination in the viewing environment. When the eyes are exposed to low levels of illumination, the visual receptors compensate by becoming more sensitive. When this happens, the apparent luminance and colorfulness of an image is lowered.

Observer local-brightness adaptation: When looking across a room and out a bright window, our eyes can quickly adapt to different levels of luminance. Detail can be seen in the shadows in the corner of the room, and in the brightest clouds in the sky. This is why the human visual system is able to process HDR (high dynamic range) images. This local adaptation cannot occur in non-HDR imaging systems. There is a set exposure range that it can capture and display. The roll of the tonescale for shadows and highlights is believed to be preferred because of this type of adaptation.

This psychophysical signal processing is independent of colorimetry. The colorimetry describes the color stimulus that meets the eye. That does not mean that imaging systems cannot be based on colorimetry. Colorimetry is a perfectly acceptable form of encoding, as long as the medium is not changing. For example, if a copier is being designed to make copies of reflection

prints, measuring colorimetry and reproducing that colorimetry on another reflection print is perfectly appropriate. This is only possible because the reproduction is being made of a reproduction. Their colorimetric characteristics and expected viewing conditions are the same. However, if a slide film was scanned and colorimetrically encoded for print, it would look dark, high in contrast, and tinted towards cyan. This is caused by the fact that the slide would be projected in a dark environment with an observer who is chromatically adapted to a different white point due to the characteristics of the projector lamp. These viewing condition differences change the perceived contrast and color balance of the image. The colorimetric characteristics that produce a pleasing image will differ from a reflection print, since a print would likely be viewed in an average surround under a different type of illumination.

Therefore, these different conditions must be accounted for when dealing with images from different input mediums. These colorimetric differences can be accounted for with appropriate mathematical transforms. When an imaging system such as film and video post-production has multiple input and output mediums, the input and output signal processing transforms must be separated to avoid creating transforms for *each* combination of input and output. The system, then, must have a unified input/output interfact at the center of the system where there is a standardized representation of color. This concept is referred to as a *color encoding specification*.

2.2 Imaging Fundamentals

2.2.1 Tonescale

Given the importance of luminance information to the human visual system, tonescale is the most important aspect of image reproduction. *Tonescale* is a term used to describe how a camera or piece of film renders the range of light information that it is able to capture. It is essentially a plot from physical amount of light to the digital code values or film densities that are produced. A linear tonescale would result in a straight-lined reproduction of the captured scene. While this sounds good on paper, the image will appear to be low in contrast and lack vibrant color: a flat and dull reproduction. This is caused by all of the adaptations and color appearance changes that occur between the real scene and the reproduction viewing conditions.

For this reason, the contrast is usually raised by increasing the slope of the tonescale in the middle range of luminance where most of a properly exposed scene will lie, such as the subject's face. With this steeper slope to the tonescale, the shadows and highlights or "toe and shoulder" will quickly reach the limits of possible digital code values or film densities. Therefore, the toe and shoulder are usually rolled off to compress the shadows and highlights, but maintain some detail by squeezing them into what remain of the digital code values available. This creates an "S-shaped" tonescale curve that is common to film and many other image capture mediums.

This type of rendering is not a novel idea. Painters discovered this concept hundreds of years ago, when paint became rich enough to reflect a high dynamic range of light in a scene. They began exaggerating the contrast of the subject's faces, while compressing the highlights of bright windows and shadows. Evaluating the tonescale of hundreds of classical paintings revealed their tonescale and simultaneous contrast to be very similar to the standard for HDTV broadcast today. When thinking forward to a new standard for digital motion picture rendering, it only makes sense to begin development where hundreds of years of work have left off.

2.2.2 Color Reproduction

Film renders skin tones slightly warmer than they appear in real life. This is an aesthetic decision based on feedback from viewers who seem to prefer it. This shift is an example of a color reproduction change that is also possible in a digital space. Just as tonescale is used to shift the overall contrast and appearance of an image, the colors can be similarly shifted. There are numerous techniques for changing the color reproduction of an image. Some readers may be familiar with their favorite program, which allows them to “color balance” an image by scaling the red, green, or blue channels, or even shift the “color temperature” from cool to warm. These are front-end controls for the complex image processing techniques going on behind the scenes.

One technique for modifying color reproduction is to change the individual tonescales for the red, green, or blue channels independently from one another, making the shadows, midtones, or highlights shift towards certain colors. Another technique is to employ a 3x3 matrix. Since there are three channels (RGB) for each pixel, a simple matrix-vector multiplication will shift the color of a pixel one way or the other.

2.2.3 Color Grading vs. Color Management

There is a distinct difference between *color grading* and *color management*. The techniques described above are common to both, but the stage at which they are done is important. A camera engineer makes certain decisions as to how they want their images to look *directly out of the camera*. A colorist will then sit down in front of a color corrector and tweak the image until the director or D.P. is content with it. The techniques and image processing fundamentals described above are common to both of these stages.

The problem arises when images from different cameras are intercut or composited together. Film and video for example differ entirely in their reproduction of scenes. The colors and tonescales will not match. In addition, transmitted video signals often sacrifice color information for bandwidth efficiency.⁷ This is currently a major problem, and one that the Academy hopes to eliminate with the completion and adoption of the IIF. The colorist and directors waste time and money tweaking images to the point where they often have to say “they’re not going to notice that.”

“Every successful color-imaging system employs one or more means for controlling and adjusting color throughout the system.” -E. Giorgianni

Proper color management requires color to be encoded specifically and unambiguously. This requires a color encoding specification to be defined so that a color encoded can be perfectly reproduced, regardless of where it came from. Another requirement is that all input and output mediums are characterized. It must be known what color will be reproduced if a certain scene color is captured, as well as how that reproduced color will appear to the viewer on the output device. Once these characteristics are known, through a characterization process, the camera values can be transformed into specifically encoded colors, which can be transformed to display, as intended, on the output device.

It is important to note that even in a properly color-managed workflow, color grading should be the *last* stage of post-production. When images come into post-production, they will be in the common encoding space. This way, the rough cut can be done and special effects/CGI elements can be composited seamlessly. Once these elements are combined, color grading should be done. The point is to avoid color grading different material and then splicing them together. Often,

material from different encoding spaces will differ so much that even extensive color correction cannot match them.

Therefore, the most important component of a proper digital color management system is the common encoding, or color encoding specification, which will be discussed further.

2.2.4 Standards

Imaging system standards have been around for decades and provide for an established paradigm for capture and workflow. As mentioned, Kodak was at the forefront of the digital intermediate during the 1990s. Their products were ahead of their time. Cineon was adopted into the DPX reference printing density encoding standard that still flourishes to this day¹¹.

Right now, digital intermediate images are encoded and evaluated using printing-density (PD) values usually in the form of DPX files. These printing density values are useful because they relate to the density that the print film will ultimately “see” when there is a final output to print film. Knowing the characteristics of a standard print film, such as Eastman Color Print film, and those of a standard film projector, print film emulation techniques are used in the color correction suites to give the director, D.P., or colorist a realistic look at what the final image will look like in theatres. For this reason, new print film stocks are only released every 10-15 years. If a new print film stock were released, the print film emulation in D.I. would no longer be accurate, and a new characterization would be needed. If new negative stocks are released, there is no problem, as they can always be printed to the same print film and can therefore be encoded with the same printing-density values. Regardless of these mastering standards for digital intermediate work, the encoding was based on film. The subtractive and logarithmic nature of printing density creates a problem when working with images on video monitors. Video is an additive system consisting of gamma-encoded linear exposures.

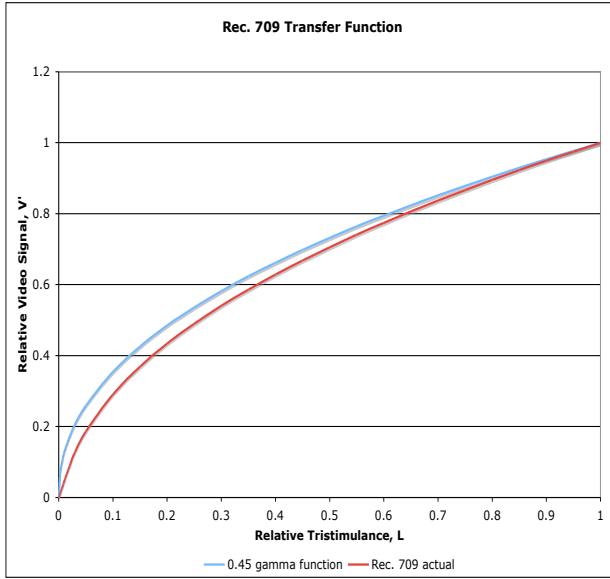
The Cineon system solved this issue by establishing a “log-to-lin” process which could convert data (DPX) to video, and vice versa. It was a simple approach to the problem, but it worked. Print film emulation transforms were then developed to be able to give a colorist the best approximation of what the image will look like once it is printed back to film and projected in a theatre. This is essential in a digital intermediate facility where color-critical decisions are made every day.

One of particular importance to this paper is the *ITU-R Recommendation BT.709* standard for high-definition television. Simply referred to as Rec. 709, this standard for the capture and display of high-definition video was established in 1990.¹² As mentioned earlier, it is based on decades of research and development of the “preferred” reproduction of a given scene. Compared to the “film look,” it is a more faithful rendition of what we see with our own eyes. This is due to the nature of television imagery versus that in the movie theater.

A video standard such as Rec. 709 must contain at least these three necessary components:¹²

1. **Primary Chromaticity:** the CIE chromaticities (see Section 2.1.1) for the display primaries of red, green, and blue
2. **Phosphor Matrix:** a 3x3 matrix derived from the primary chromaticities and the white point chromaticity relating amounts of the primaries to the actual scene XYZ tristimulus values (see Section 2.1.1)
3. **Non-linear Transfer Function:** the equation for the mapping of linear light to digital code value at capture (i.e. tonescale) and/or at display (see Rec. 709 transfer function in Figure 5).

With the establishment of the Rec. 709 standard, production and post-production facilities are able to conform to a standard with which they can properly view and manage images, independent of the capture or display device. It also allows for the older NTSC/PAL standards to be inherently compatible, so that it can be more widely adopted. In a closed and calibrated workflow, this works extremely well. SoFA at RIT is not closed and calibrated to any extent. The cameras available have not been characterized, and the workflow that follows is not calibrated. In fact, the Apple Cinema Displays in the labs do not offer any in-monitor controls for calibration purposes. This lab is also where the Digital Color Correction course is taught. While the professor took note of the discrepancies that were inevitable between each station, it would be more valuable if students could be more confident with their decisions.



(a)

$$V'_{709} = \begin{cases} 4.5L & 0 \leq L \leq 0.018 \\ 1.099L^{0.45} - 0.099 & 0.018 \leq L \leq 1 \end{cases}$$

(b)

Fig. 5: The ITU-R Rec. 709 transfer function for encoding linear scene luminance (L) into HDTV camera signal (V').

With digital projection on the forefront, the motion picture industry's biggest seven studios joined forces to establish the Digital Cinema Initiatives (DCI).¹² The standards they have set up are essentially the building blocks needed for *after* post-production. Once a film is truly done and a look is finalized in post, DCI standards dictate the format and compression that is accepted by theatres across the country for digital exhibition.

Implementation of the proposed Academy framework would be valuable to an institution such as SoFA. Not only would the creative side benefit from the image consistency, but digital cinema students would learn from and work at maintaining the closely calibrated system.

2.3 Color Encoding Specification

A color encoding specification (CES) is a necessary description of the method and metric by which colors are encoded in a digital color management system. A system with an unambiguous and

strictly defined color encoding specification allows for much more system flexibility.⁵

The “metric” of an encoded color, the color space and numerical representation of the color, is an obvious necessity for any color imaging system. This could be CIE XYZ, CIELAB, CIELUV, or RGB. However, the “method” for a color encoding specification is a component often overlooked. The method describes the actual meaning of the numbers encoded. There are many different options for which method of color encoding to use. Most often the system’s input and output devices and mediums will dictate the best choice. If a system is producing high-quality slide films, a smart choice would be to encode the colorimetry of the reproduced slide film when projected.

In addition to the colorimetry of a certain medium being encoded, such as projected slide film, there needs to be a set of encoding reference viewing conditions. Without this, the color appearance would change drastically if the colorimetry were viewed in a different environment, making the encoded colorimetry ambiguous. With colorimetry and a set of reference viewing conditions, there is no ambiguity. If a color is shown with a set of encoded colorimetric values in the encoding reference viewing conditions, the color will appear exactly as it was intended and encoded.

For the purposes of film and video post-production, the situation is more complicated. There are many different types of input mediums, ranging from transmissive media such as motion picture and slide film, additive devices such as digital cameras, as well as computer-generated imagery (CGI) elements. The vast amount of information these devices are capable of producing must be preserved for post-production. In addition, the outputs range from standard definition DVD’s to high quality print film recorders, and the final decision on the appearance of the image must be preserved on each output medium.

The Academy committee spent a significant amount of time trying to figure out a color encoding specification that would allow for all input mediums to be encoded, intercut, and composited, but also maintain consistency through the output devices. The conclusion was that there needed to be two separate color-encoding specifications one for input (ICES) and one for output (OCES). There are two main reasons for this.

The first is that the amount of information encoded by an input medium such as motion picture negative film would be greater than that of the output mediums. If an encoding method was chosen based on the output colorimetry, information would be lost. In high-quality digital post-production, this information is necessary in order to pull information out of shadows and highlights in color correction. The second reason is that if the colors are encoded based on the input color encoding specification, the output devices will have too much control over the final look of the image, since it is still in its “original” state. The director or other creative controller of the production will decide on a certain “look” and that is what they will want to see emulated on each output device.

The ICES and OCES approach is a perfect solution to handle a complex system like that of motion picture post-production. The ICES values are encoded in terms of the input scene colorimetry which stores an extreme amount of image information. In this “scene state” is where color correction, compositing, and all other post-production techniques are applied. Once the image is ready for output, it is transformed into OCES, or a “rendered state.” In this state, the image is final. The director has decided on a “look” is the one that should be emulated from OCES onto all output mediums.

2.3.1 Academy Color Encoding Space

The approach used for defining the ICES is “scene-based.” This attempts to predict the original scene colorimetry based on the characteristics of the input devices. This is the *only* approach that will allow for all input mediums be encoded with no loss in information and to be seamlessly encoded in this type of system. This is possible because a scene-based approach can achieve complete input compatibility by relating the input devices to original scene colorimetry and eliminating any disparities between the input mediums. In addition, the approach is not based on rendered or reproduced colors, so it allows for the preservation of all of the captured scene information.

Achieving complete input compatibility comes at a cost. With this type of compatibility, the differences between input mediums is completely removed and colors are encoded based on actual scene colorimetry. This is useful for applications where the color accuracy is the goal such as medical or military applications. Removing these input medium differences also removes their inherent characteristics or “look,” which may be preferred. In other words, this would remove any of the color alterations that may have been desired by the director who chose the input medium based on their creative intent. The use of a *universal input transforms* rather than *product specific input transforms* will preserve the preferred “look” of the different input mediums, at the cost of input compatibility.

A common misconception of scene-based encoding methods is that the encoded values are only scene-based if they accurately represent the color appearance of the scene. As in the example above, the characteristics of an input medium are often preserved into scene-based encoding. These encoded values are still scene-based because they represent the colorimetry *of the scene* as seen by the specific input medium. Another example would be if a scene is brightly lit and then dimly lit. The colors will be brighter and more saturated when brightly lit. Both situations would be encoding of the original scene, even though the scene-based encoded values would not match.⁵ Similarly, if a camera is known for its more saturated colors, the scene-based encoded values would be more saturated, but still based on the colorimetric characteristics of the *scene* rather than a *reproduction* of the scene. Scene-based encoded values, therefore, can change depending on variables such as film stock, exposure decisions, and lighting. The key is that the colors are in “scene state” rather than “rendered” or “reproduced state.” This concept should seem natural to someone who understands film. The negative is essentially in scene state until it is rendered when printed.

This approach of scene-based encoding for an Input Color Encoding Specification was used in the Academy’s document detailing ACES: Academy Color Encoding Specification. In order to make the scene-based approach work, a specific encoding method was needed. The Academy committee decided to base the encoding on a *reference image capture device* or RICD. This is a hypothetical device with spectral sensitivities that are CIE color-matching functions and primaries that encompass the entire spectrum locus of physically realizable light. Since it is a hypothetical device, there is no limitation on the dynamic range or color gamut. Any color that the human observer can be represented. As needed for a strict representation of color, ACES also specifies a set of reference input-encoding conditions, representative of a typical daylight-illuminated outdoor scene:³

0% Flare: Any flare light is considered part of the scene itself

Normal surround: Objects surrounding colors are all similarly illuminated.

Luminance level: $> 1600 \frac{cd}{m^2}$

Observer adaptive white: $x = 0.32168, y = 0.33767$
(D60 chromaticity and ACES unity RGB)

If this hypothetical RICD camera were set up in front of a scene under the reference input-encoding conditions, the camera would directly produce ACES values.

Through the use of *Input Device Transforms*, any input medium can be converted to ACES values, the scene-based encoded values that would have been produced had the RICD captured the same scene.

2.4 Input Device Transforms

An input device transform (IDT) converts code values from a given image capture device, in this case the HVX, to ACES RGB relative exposure values. These ACES RGB values approximate the colorimetry that would be obtained if the scene were captured by the Reference Image Capture Device.

The accuracy of an IDT to predict these exposures depends on:⁴

1. The degree to which the image capture device's response to light is colorimetric (i.e. its similarity to a linear combination of color-matching functions)
2. The difference between training spectra used for IDT computation and the spectra of the actual scene itself.
3. The choice of distribution of color errors in relation to scene colors.

Because the RICD is a linear, colorimetric camera, the IDT must account for the tonescale and colorimetric characteristics of the camera. Therefore, the IDT accounts for the signal processing that occurs inside of the HVX prior to output code value formation.

The IDT also compensates for the difference between scene adopted white chromaticity and the ACES neutral chromaticity (D60). The “scene adopted white” is the spectral power distribution in the scene that, as seen by an image capture or measurement device, is considered to be perfectly achromatic. The most common way of dealing with this difference is a white balance. For video, a simple white balance will force the scene white to equal $R=G=B$, which will appear white on the monitor that it is shown on.

However, there is another method of compensation, known as *Chromatic Adaptation*, designed to better approximate the colorimetric difference in terms of the human visual system, rather than a simple white balance which just adjusts the RGB signals to achieve equal $R=G=B$. This method will be known as Method 1. The white-balance approach will be known as Method 2.

Method 1: Chromatic Adaptation: The first approach is a method of using a chromatic adaptation transform (using Von Kries method with CAT02 matrix⁸) to convert the scene XYZ tristimulus values to what they would have looked like had the observer been chromatically adapted to a D60 illuminant. These new XYZ values become the aim of the IDT, i.e. it will convert camera linear code values to chromatically-adapted-to-D60 XYZ values, which can easily be encoded as ACES RGB values.

Method 2: White Balance Only: The second method is to apply a simple white balance gain factor to each channel so that the camera RGB output for neutrals under the scene adopted white become ACES R=G=B, which will ultimately display as a neutral D60 when displayed as ACES.

Each choice of scene adopted white will be a unique IDT. This is similar to the standard of film stocks as being either tungsten-balanced or daylight-balanced. In addition, each method will produce its own IDT for each scene adopted white, and it will be interesting to see the colorimetric consequences of each method. Both of these methods will be investigated, and three real illuminants for scene adopted white will be examined: studio-tungsten lights, HMI (effectively daylight), and KinoFlo Daylight (fluorescent daylight).

The core of an IDT should be a 3x3 matrix, and there are several advantages to this. The advantages of this are that a 3x3 matrix fits into current industry practice closely, so implementation is simple. Forced row-unity will preserve neutral code values from camera RGB to ACES RGB, and vice versa. It is simple to understand mathematically and is easily invertible.

An input device transform for the Panasonic HVX-200 has two parts. The first involves a one-dimensional look-up-table (1D LUT) which accounts for the non-linearity of the signal that comes out of the camera. Since CIE colorimetry is in a linear space, the gamma power function that relates the linear luminance of the scene to the output camera video signal must be inverted to predict the linear scene luminance levels. Once the camera values are linearly related to light, the second part, the 3x3 matrix, accounts for the colorimetric differences between scene-to-RICD and scene-to-HVX colorimetry.

Once in ACES space, there is likely to be some margin of difference. The developer of an IDT is responsible for coming up with a metric on which to optimize on, since a simple ΔRGB would be non-uniform. For the purposes of this project, $\Delta E 2000$ will be used as an optimization metric.¹³

2.5 RRTs, OCES, and ODTs

Just as every image capture device has certain colorimetric characteristics which affect the overall “look” of its imagery, there needs to be a decision made on how to *render* the scene once the ACES values are derived. Take a bright outdoor scene for example. On a clear day, outdoor luminance levels are typically from 10,000-25,000 $\frac{cd}{m^2}$, while a typical theatre luminance level is around 48 $\frac{cd}{m^2}$. This 500:1 difference in overall brightness, along with other factors discussed in Section 2.1.2, requires an image of the outdoor scene projected in a theatre to have increased contrast and saturation in order to produce a pleasing *reproduction* for the viewer. Even if they have never *been* to the location in the scene, they will have expectations, memories, and preferences as to what they prefer to see in that type of scene.

Since ACES is a linear scene-based color space, it is not ready for viewing in that state. It must be rendered somehow to a viewing or rendered state. This can be thought of as the equivalent of the film printing process. A large range of scene exposure dynamic range is stored on a motion picture film negative (although not for direct viewing) and the *reference rendering transform* (RRT) is the equivalent of the film being *printed* or *rendered* onto the positive-looking print film. The dynamic range of print film is not as high as that of negative film, so a certain exposure range must be chosen when printing.

This similarity to the film process is not a coincidence. The Academy had to design a system for digital motion picture mastering that did not limit the amount of information at hand during the post-production or digital intermediate stage (working with images in their ACES or scene state). However, there needed to be a clearly-defined RRT so that when a director or D.P. decides on the final look of their film in the color correction suite, they will be looking at it through *the* RRT that will be the standard rendering for mastering from there forward.

The encoding reference conditions is the key to communicating color. As explained earlier, colorimetry is not enough to describe color appearance. The viewing conditions have a profound effect. If a stimulus is sent to from one person to another, even on the same type of screen, they would have no idea if the color was properly communicated. In OCES space, the image is rendered to be viewed on a hypothetical reference projector in the dark-surround viewing conditions. While this projector doesn't exist, a simple ODT to fit the OCES values onto a real device is used to be able to produce the color that is intended by the OCES values.

3 HVX CHARACTERIZATION

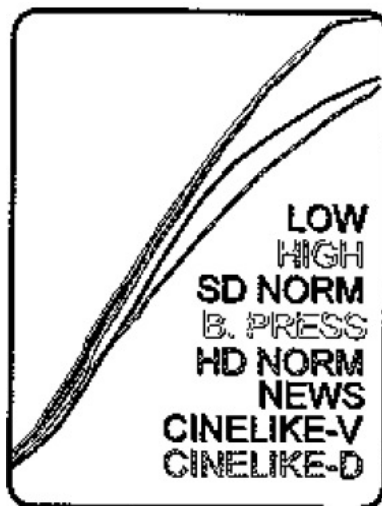
3.1 Grayscale Response

Because the input device transform matrix operates in a linear space, the input code values to the transform should be radiometrically linear code values from the given image capture device. A “radiometrically linear code value” is an encoded value that is linearly related to the physical amount of light, or luminance at that pixel. Devices like professional digital cameras make this fairly straightforward by outputting RAW linear code values that can be direct input for an IDT. However, to investigate a high-definition video camera like the HVX is more challenging. High-definition video cameras typically output 8-bit video signals: code values ranging from 0-255. At low code values, the human eye is very sensitive to small changes in luminance. If the difference in luminance between adjacent code values is noticeable to the eye, a smooth gradient may appear contoured. To avoid this undesired “contouring effect,” a *gamma* is applied to the linear values in order to make effective perceptual use of a limited number of bits per pixel.⁷ Developing an input device transform requires extensive testing to investigate these characteristics.

Determining the grayscale characteristic of the HVX is necessary to be able to linearize the camera's output code values for use during IDT computation. To do this, a Kodak Q14 Color Separation Guide was used. The target was developed as a quality control device for comparing tone values of reflection copy with its reproduction. It contains 20 grayscale patches in 0.10 density increments between 0.0 (white) and a practical printing black of 1.9 density. For our purposes, it would allow us to compare the camera's output code values with physical amount of light coming off of each patch to plot the tone reproduction curve of the camera.

The HVX has several on-camera settings that affect this curve. The first is gamma, which applies a non-linear look-up table (LUT) to the linear signal coming off of the camera. With the ultimate goal of reversing this LUT, the gamma decision is very important. Therefore, the grayscale characterization experiment was done with all possible gamma settings on the HVX: HIGH, MED, LOW, HD NORM, SD NORM, CINE D, CINE V, and NEWS. After researching, the curve set shown in Figure 6 is the best approximation of the curves found. Not only are the curves hard to identify from each other, but there are no labels on the axes. Clearly, there is a lack of detail needed to be able to make any technical decisions based on it.

Fig. 6: The grayscale response curves shown in “A Guide to the HVX-200 Camera” published by Fiercely Independent Films, Inc., 2007.



The HVX tone curve is also affected by the choice of “Knee” setting in the camera. The knee determines at which code value to flatten out the tone curve in the highlight region, in order to control compression and clipping of highlights. The settings of HIGH, MED, and LOW are available. The Panasonic manual’s lack of detail in explaining these settings raises the question: does a HIGH knee setting mean a) a high “amount of knee,” which would mean engaging at a low code value, or b) a high code value that the knee engages at. Because of this ambiguity, all knee settings were tested with each gamma setting.

The goal of the following experiment was to find the gamma and knee setting that gave a response curve that would mathematically simple. This gamma response can then be incorporated into a mathematical model that will linearize the signal that comes out of the camera. Optimally, this would be a simple power function of some value. The ITU-R Rec. 709 standard HDTV transfer function (Figure 5) contains a gamma power of effectively 0.5. If this encoding is sent to a monitor with a gamma of 2.5, this will show an effective gamma on the image of 1.25, which is preferred in a video reproduction. The SMPTE standard for video mastering calls for a 10% surround luminance. For a dim surround environment, as discussed before, the perceived luminance contrast is lowered, so the raised gamma (contrast) of the final image is necessary to produce a pleasing image. It was expected that one of the gamma settings on the HVX would produce a curve that resembled a simple power function in the neighborhood of 0.5, similar to that of the standard HDTV transfer function of ITU-R Rec. 709 (Figure 5).

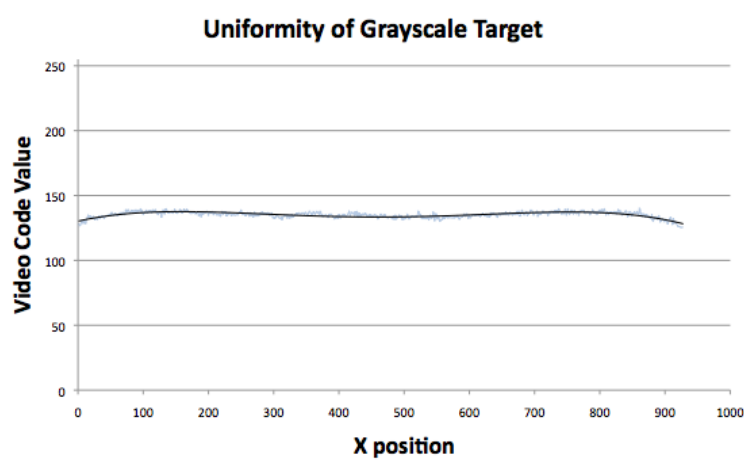
3.1.1 Experimental Setup & Procedure

The target was set up in front of the HVX and lit with studio-typical tungsten lighting. An incident light meter was used to adjust the lighting in order to approximate uniform lighting across the target. This was achieved to a degree of 2 cd/m^2 , with a slight raise in luminance towards the darker patches, which was acceptable, but must be considered as uncertainty in subsequent analysis. The camera was set up on a tripod in order to capture the target across the entire frame. From beside the camera, the light meter was used as a spot meter to measure the actual illuminance for each patch.

These measurements were made in order to be able to calculate actual scene exposures for each patch, using a photographic formula which involves illuminance (i.e. 60 cd/m^2) and f-number (i.e. $f/5.6$) to calculate the physical amount of light entering the camera (i.e. lux-sec). An example of an image captured using this setup is shown in Figure 8.

Uniform lighting upon the target was important, and the setup was carefully lit to account for this. Even so, achieving perfectly uniform light with studio tungsten lights is a challenge. As explained above, the photographer's incident meter was used as a guide for setting up the light. The actual test for uniformity was done after the experiment, and involved analyzing a row of pixels in the upper gray areas of the chart just above the patches (as you can see between the patches and the numbers in Figure 8). The results of this analysis are shown in Figure 7. A smooth fit of the otherwise noisy code values was made to show the overall shape of the lighting distribution. The graph's "double hump" is evidence of the two-light setup used. The data showed a variance of 14.7 CV. This was acceptable, since the primary concern was the illuminance readings from each patch.

Fig. 7: Plotted are the code values of a row of pixels spanning the gray area just above the patches, to test for lighting uniformity.



The target was shot with each combination of GAMMA and KNEE setting that the camera is capable of. With CINE D and CINE V settings, the KNEE is locked, implying that these gamma settings have their own corresponding KNEE setting that cannot be changed. With each combination, the target was shot at an iris setting of $f/2.8$, $f/5.6$ and $f/11$. This would allow for the signal to be pushed into clipping to 0% in the blacks with the $f/11$ setting, and to 100% in the whites with the $f/2.8$ setting.

3.1.2 Results

These exposures produced three separate response curves for each GAMMA + KNEE combination. The assumption could be made that the curves, when related back to scene exposure in lux-sec based on the $f/\#$ used, would line up with each other. This turned out to be a false assumption. The placement of the curves in relation to the horizontal exposure axis was *different* for each exposure and gamma settings. This can be seen in the curve shown in Figure 9.

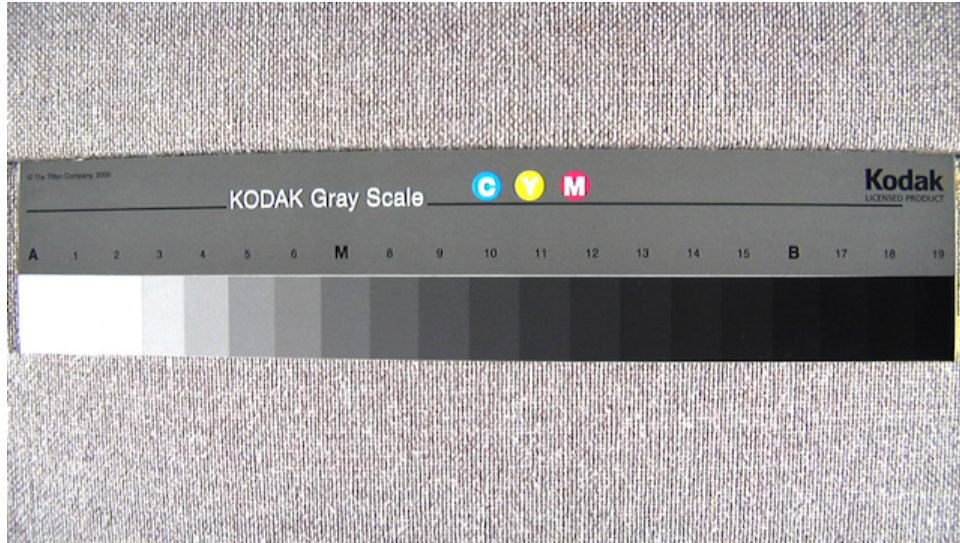


Fig. 8: An example of an image captured during the grayscale experiment. This was captured with a HIGH knee and HD NORM gamma, at an f-stop of 2.

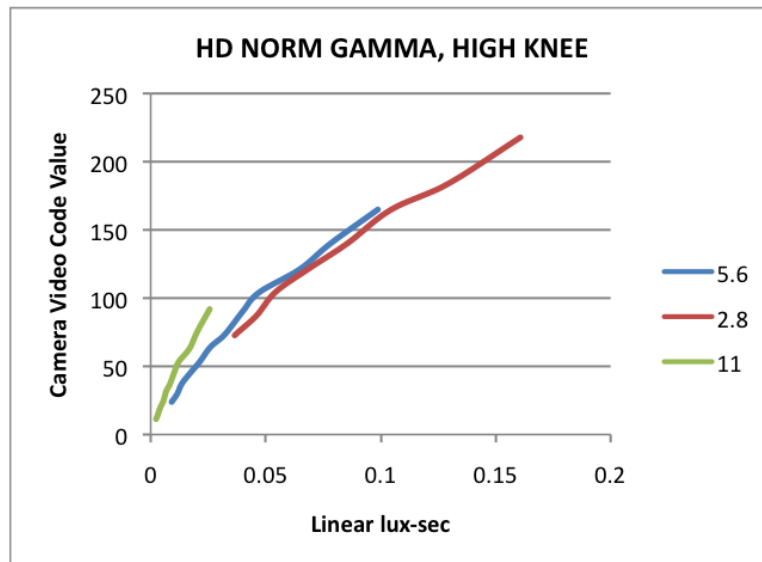


Fig. 9: The effect of the HVX “nearest neighbor” approximation of f/#, caused by a smooth iris control rather than a rigid one.

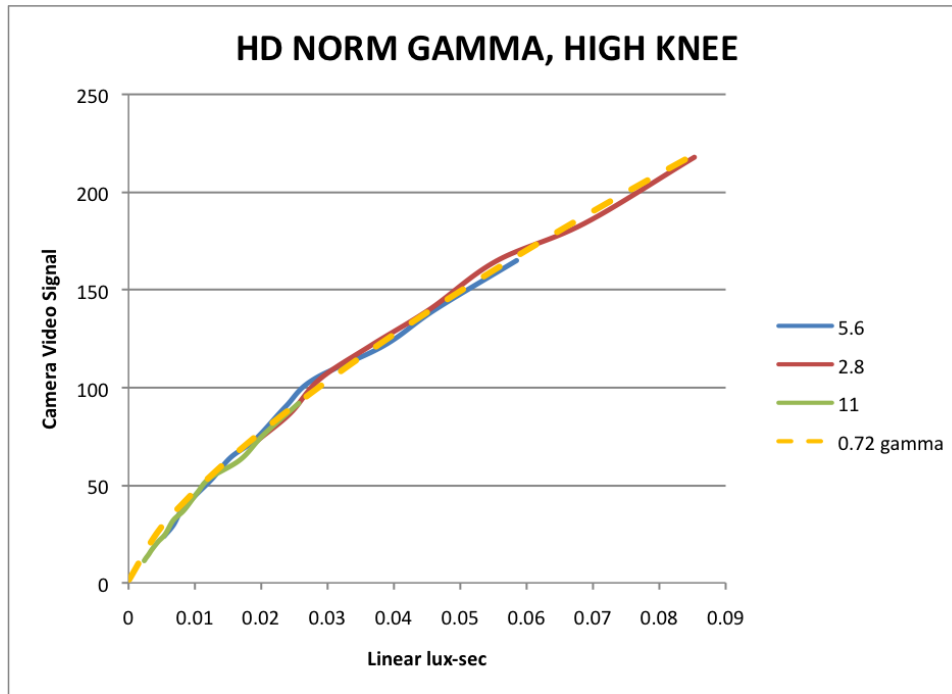


Fig. 10: The HD NORM with HIGH KNEE response curve (after shifting), with a best-fit power function of 0.72.

It was decided that the cause of this discrepancy was due to the optical behavior of the camera's iris setting. The iris wheel does not “click” to each f/# as do some professional motion picture cameras. The iris wheel, instead, is a smooth control of the aperture opening, and the digital f/# displayed is its closest f-stop. This is likely a decision that was made by Panasonic to give finer control of exposure, at the cost of the precision needed for these kinds of technical investigations.

For this reason, it was necessary to shift the curves to line up. For the purposes of finding the gamma setting that was the simplest mathematically, the primary concern is the shape of the curve, rather than the placement along the exposure axis. The f/11 was decided to be the “anchor point” to which the other curves were shifted to make one smooth curve containing all three response curves.

These curves gave evidence that the HIGH setting for KNEE was desired for the purposes of this project. The reason for this is that a LOW setting rolls off the highlights at approximately a 80% signal of the scene to prevent overexposure of highlights. This effectively flattens the curve out to a straight line from 80-100%. The HIGH setting does not flatten out, giving the response a smoother curve and closer fit to a simple power function.

After investigation of the response curves of the different GAMMA settings with a KNEE setting of HIGH, the simplest curve that fit best with a power function was HD NORM. Being that HD NORM is the “normal” choice to produce HD-like images, the decision was made to continue the model with **HD NORM GAMMA** with **HIGH KNEE** setting. The effect of fitting this response curve to a power function is shown in Figure 10.

This decision would allow for the linearization of the video code values that come out of the camera, which could predict linear scene exposures. This is a necessary step in developing in IDT,

because of the fact that the input code values must be radiometrically linear to be able to produce ACES RGB linear scene exposures.

3.1.3 Instrument Flare

The Kodak Q14 grayscale target was measured using a photographer's spot meter to measure the illuminance at each patch. However, at the darkest end of the scale, these measurements flattened out. This suggests some amount of flare within the measurement system, preventing it from distinguishing between low levels of reflected light. This created kinks in the luminance plots that needed to be chopped off in order to only use the parts that were accurate.

To investigate this further, the Q14 target was measured using a Color Eye 700 instrument to measure spectral reflectance. This instrument was built for this kind of task, so the flare was assumed to be much lower than the photographer's light meter. The instrument indeed gave more accurate readings of spectral reflectance for each patch.

However, the data provided by Kodak as part of the Q14 package did not match either the light meter or Color Eye 7000 readings. The origin of Kodak's data is a densitometer. This problem raises the question: How accurate can a professional or semi-professional be, with the equipment at hand? The photographer's light meter is a poor man's approach to this type of experiment, but was the only instrument at hand. The price tag for the Color Eye 7000 is in the thousands, and the system still had inherent flare preventing the measurements of dark patches from being perfectly accurate.

3.2 Spectral Responsivity

The next stage of the HVX characterization process was to determine the effective spectral responsivity of the CCD sensors inside the camera. Camera manufacturers often provide the spectral sensitivities of their digital cameras or film stocks to customers. The data provided, however, is gathered by exposing the actual sensor to narrow wavelengths of light, rather than the camera itself. Therefore, the manufacturer data is not an accurate prediction of the *effective* spectral response taking into account the effects of the optics and signal processing of the camera. Panasonic does not provide this data for the HVX-200A, likely because the majority of its customer base are not engineers who require this knowledge. Therefore, the camera needed to be exposed in a way that would predict the exposures for each wavelength.

3.2.1 Experimental Setup & Procedure

The Munsell Color Science Laboratory @ RIT has a bench containing a set of instruments specifically designed for determining spectral sensitivity of digital cameras. Numerous digital still cameras have been mounted on the bench with successful results. There is a tungsten bulb that is controlled by 10 nanometer slits to let only narrow spectral bands of light through the system. These 10 nm sections of the spectrum are shined into an integrating sphere in order to create a completely uniform area of light which exits the sphere towards an optical bench. These components combine to form what will from now on be called a *monochromator*.

On the optical bench, there is a Photo Research PR-704 telespectroradiometer which can measure, with great precision and high signal-to-noise ratio, the spectral radiance of the light exiting

the integrating sphere. The telespectoradiometer is pointed into the integrating sphere and connected to a workstation that takes measurements of radiance ($W/sr/m^2$) at 4nm increments from 380-780 nm.

The HVX-200A was the first video camera to ever be mounted on the instrument. The camera was mounted, directly beside the PR-704, and the lens was zoomed to nearly fill the frame in the viewfinder. The focus was adjusted to produce an *almost-in-focus* image in order to create a more uniform, noise-free image of the monochromator. The settings below were used, and are typical of those used by students of SoFA for productions:

GAMMA: HD NORM

KNEE: HIGH

FORMAT: DVCPRO HD 720p 24pN

MATRIX: NORM

WHITE BALANCE: 3.2K PRESET

For each 10nm increment of the monochromator light, the camera recorded 1-2 seconds of the monochromator. Meanwhile, the PR-750 took measurements of radiance for each 10nm increment. The spectral radiance curves shown on the workstation were consistent with the design of the monochromator to produce narrow spikes of light with a half-peak bandwidth of 10 nm. This was repeated for each 10nm increment between 400nm and 700nm to produce a total of 30 clips on the HVX and 30 corresponding radiance measurements on the workstation.

3.2.2 Results

The images recorded at each 10nm increment by the HVX were extracted using still-image TIFF output from Final Cut Pro. To confirm this image extraction did not affect the video code values, the images extracted were re-ingested into Final Cut Pro and the histograms were lined up to confirm a match. An IDL program was written to output, for a specified area within the light of the monochromator in each frame, the *mean RGB code values*.

After obtaining the R'G'B' (nonlinear) code values (CV) for each of the 10nm increments, these code values were linearized using the *inverse* of the best-fit power function ($CV^{\frac{1}{0.72}}$). The subsequent *RGB linear code values* were divided by the radiance measurement made by the PR-750. This was done to normalize the code values based on the absolute amount of light that hit the center. After this normalization, the RGB numbers are essentially unit-less, but represent the “video code value per radiance” which is essentially the camera’s linear response to light. These values were plotted and are shown in Figure 11. Evidence of the “3.2K” white balance is clear, since a perfect white reflectance of a tungsten illuminant will have a spectral radiance similar to the orange line, and should produce equal RGB values off of the camera.

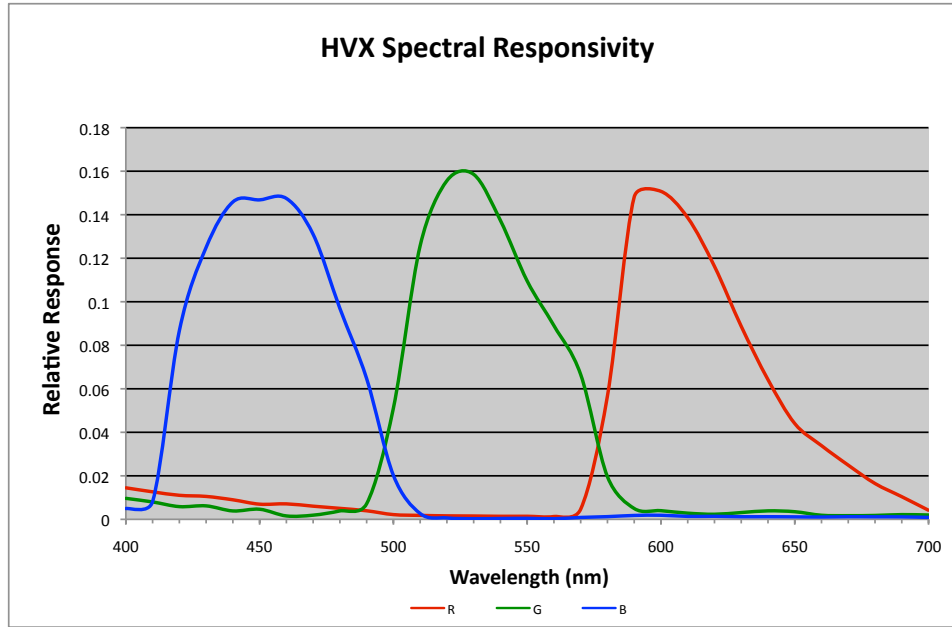


Fig. 11: The result of the monochromator experiment: HVX spectral response for red, green, and blue signals.

3.2.3 Verification

The next step was to determine the accuracy of the spectral responsivity curves that were produced using the data from the monochromator procedure. To do this, a series of patches needed to be captured by the HVX in a controlled lighting scenario, alongside a spectroradiometer capturing the spectral radiance of each patch. Cascading the spectral radiance curves of each patch with the red, green, and blue spectral responsivity curves of the HVX would produce RGB values that could then be directly compared to the RGB values that come out of the camera video signal.

To accomplish this, a GretagMacbeth ColorChecker chart was placed in a GretagMacbeth SpectraLight light-booth with the illuminant set to Illuminant A, the CIE standard for incandescent light which resembles a planckian blackbody radiator at a temperature of 2856K. The HVX camera was used to capture each patch one at a time. Meanwhile, a Photo Research PR-750 spectroradiometer was aimed at each patch individually, to measure the radiance from wavelengths of 380 to 780 nm, at 4 nm increments. An image from the HVX is shown in Figure 12.

To extract the “camera RGB” values from the HVX, the video was first imported into Final Cut Pro. Quicktime Conversion was then used to export a TIFF still-image of each patch, and the code values for each patch were averaged using an IDL function to give the mean RGB values for each patch. In Excel, the spectral data from each patch was cascaded with the respective RGB spectral responsivity curves (which uses 0.72 as gamma to calculate) to produce “calculated RGB” values. These were compared to the mean RGB values from the HVX, and the average errors between the two are shown in Table 1.

These values confirm the accuracy of the spectral response curves shown in Figure 11. The exact same camera was used with the “3.2K” white balance preset used during the monochromator procedure. Given the 8-bit video signal range of 0-255, an average error of approximately 3 CV



Fig. 12: A still image taken from the video produced by the HVX, capturing the Macbeth chart in a Spectralight booth under Illuminant A, as part of the spectral response verification process.

Table 1: The average error in code values between *camera RGB* and *calculated RGB* code values (ΔCV) of all 24 color patches.

Channel	ΔCV
R	2.96
G	2.00
B	3.20

(or 1.2% error) was small enough for the purposes of this project to continue forward. The spectral response curves could now be used as part of the IDT computation process.

Essentially, there now exists a computational “forward model” of the HVX camera. Knowing the spectral radiance of a patch (or $\text{reflectance}(\lambda) \text{ illuminant}(\lambda)$)

4 IDT COMPUTATION

Computation of an Input Device Transform requires extensive characterization of an input device. An engineer at a camera manufacturer would be the best fit for this task. They have access to the parts that make up the camera, the engineering plans that went into its creation, and likely the resources available to characterize it completely.

The purpose of this project is to determine how accurate of an IDT can be produced by a student with the resources such as those available at RIT. Using the equipment available from the Film and Video Cage at SoFA in Building 7B and the Munsell Color Science Lab in Building 18, the HVX will be fully characterized. IDTs can then be produced and their accuracy tested.

Two considerations must be taken into account when building IDTs:

1. **Spectral Sensitivity Differences:** Since the spectral response curves derived above do *not* resemble linear transformations of color-matching functions, the accuracy of an IDT for this camera cannot be perfect (see Section 2.4). A colorimetric video camera would have

response curves that can be linearly transformed with a 3x3 matrix into the CIE 1931 color-matching functions (CMFs) shown in Figure 3. Without a matrix, achieving a colorimetric camera is impossible, since the primaries needed would be too saturated to be inside the spectrum locus. This type of camera could achieve accurate color reproduction with a mathematically perfect IDT, since the RICD is also colorimetric. The conversion from one to the other would be a simple linear transform. With a camera such as the HVX, the IDT will impart some magnitude of error, simply because of the inherent difference in sensitivity curves. See Figure 13 for a visual comparison of these curves.

2. **Scene Adopted White:** The choice of a *scene adopted white* is necessary for computation of an IDT. ACES RGB values are encoded with a neutral chromaticity (D60) that may differ from the chromaticity of a neutral in the scene. As a reminder, the scene adopted white is the spectral power distribution in the scene that, as seen by an image capture or measurement device, is considered to be perfectly achromatic. IDTs must be computed for specific *scene adopted whites* and must be selected for use accordingly.

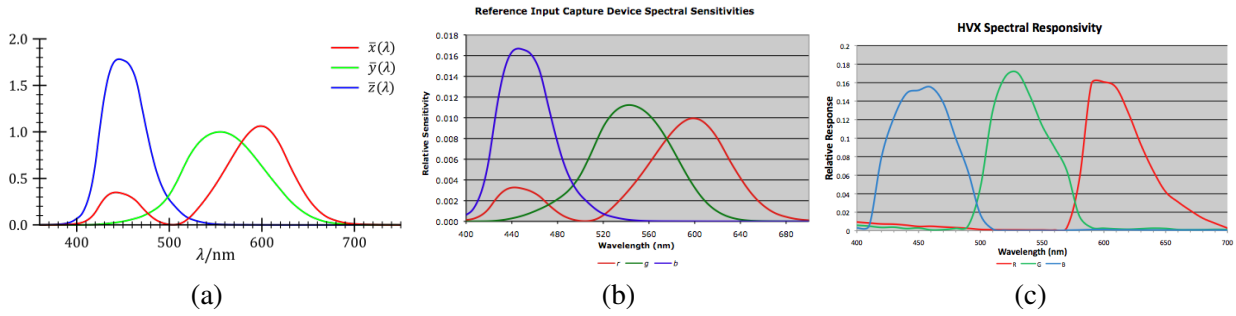


Fig. 13: It is clear by comparison that the RICD spectral sensitivity curves (b) are a linear combination of the CIE 1931 color-matching functions (a), and the HVX response curves (c) are not linear combinations of CMFs.

4.1 Methods

The computation of Input Device Transforms was done in MATLAB. The spectral response curves and non-linear transfer function for the HVX derived in the preceding sections were imported into MATLAB for use with the 190 training spectra.

The training spectra consist of 190 patches often used by engineers designing these types of transforms. The patches are used by the Academy, and were chosen as an evenly distributed spread of the entire color spectrum, including a neutral scale. In addition to spanning the color wheel, many of these reflectances were measured from real-world objects, rather than a synthetically generated set of dyes or pigments. The patches were converted to ACES and are shown as reference in Figure 14.

The first step for IDT computation was to choose typical “scene adopted whites.” Basically, that typical lighting scenarios of SOFA students had to be considered. Computationally, the easiest illuminants to begin with is a *blackbody radiator*. Max Planck determined that the SPD from a

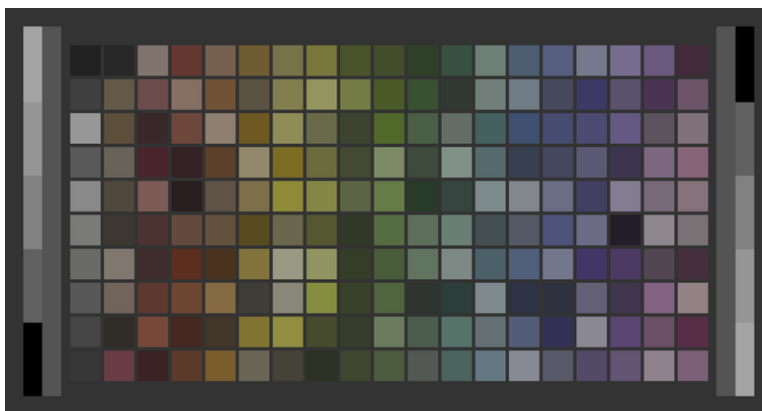


Fig. 14: A computer-generated chart showing the 190 training patches, converted from spectral reflectance to ACES values.

hot object – a blackbody radiator – is a function of the temperature to which the object is heated. Since these SPDs can be computed easily thanks to the “Planck equation,” they are an easy way to *approximate* the SPD of an actual light close that temperature. Spectral power distributions were *computed* for perfect blackbodies at temperatures 3200K and 5500K (typical of tungsten and daylight, respectively) were computed and used as a hypothetical “taking illuminant” for the scene adopted whites. It is important to note that these are entirely computational and do not represent actual experimental lighting conditions that will be used later.

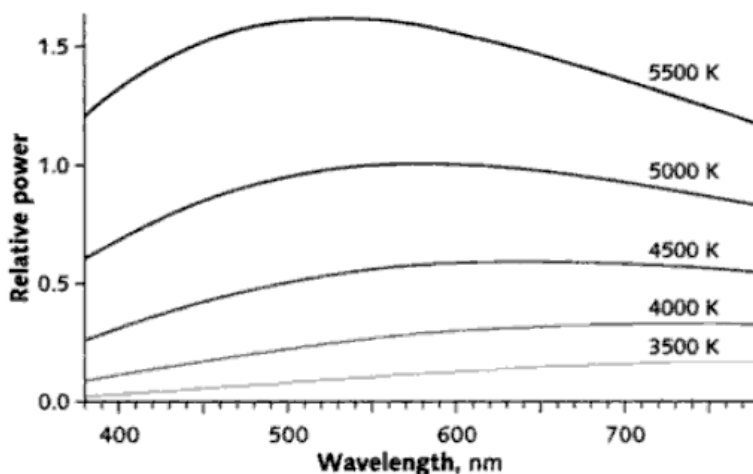


Fig. 15: A plot showing SPDs of blackbody radiators at several Kelvin temperatures.⁷

As described in Section 2.4, two approaches were taken for IDT computation. They differ in their handling of the difference between the neutral scene-adopted white and the neutral of the ACES space. The two methods and their practical consequences will now be described.

4.1.1 Method 1: Chromatic Adaptation

The first method involves chromatic adaptation of scene XYZ tristimulus values to XYZ tristimulus values that would appear the same to an observer who is chromatically adapted to a D60 illuminant. Important note:

These are *not* the XYZ tristimulus values that would have been measured under a D60 illuminant. They are the XYZ tristimulus values that would *appear*, to someone chromatically adapted to D60, to visually match the appearance of those colors to someone who was chromatically adapted to the scene adopted white.

This is an Academy-recommended option for compensation of differences between the chromaticity of the scene adopted white and the chromaticity of ACES neutral.

This technique makes sense. If a director or D.P. is on-set during production, there is a good chance that their state of chromatic adaptation is somewhere close to the adopted white of the scene. If someone is in a dim editing suite or a dark theatre, their chromatic adaptation will necessarily be close to D60, given the specifications for projector and monitor whites. The idea is that a chromatic adaptation transform will be a good approximation of what the scene would have looked like had the observer been on-set.

This idea holds true only if the camera is colorimetrically accurate. As mentioned before, colorimetric differences that a film stock imparts will come through to ACES values. The chromatic adaptation transform is based on the *human* visual state of adaptation, not the *camera*. However, since the human visual system is making the judgements on-set and in post-production to adjust the scene's look, a chromatic adaptation transform is a smart choice for compensating for the differences between the on-set conditions and encoding reference viewing conditions.

Figure 16 shows an overview diagram of the method of IDT computation using Method 1.

4.1.2 Method 2: White Balance Only

The second method uses simple white balance gain factors to produce equal RGB code values for the scene adopted white. By forcing the rows of the IDT matrix to sum to unity, equal camera RGB values will always translate to equal ACES values. These equal ACES values will display as the ACES neutral chromaticity D60. This is the standard that has been used for decades with video cameras. White balancing is mandatory during video production to ensure that the whites of the scene will become the monitor white for subsequent viewing.

Indeed, this translation of neutral RGB values will also occur using Method 1. It is important to note, however, that any color with unequal camera RGB values (i.e. any non-neutral color) will translate differently from Method 1 to Method 2. These are differences to be investigated.

Figure 17 shows an overview diagram of the method of IDT computation using Method 2.

Method #1: Chromatic Adaptation

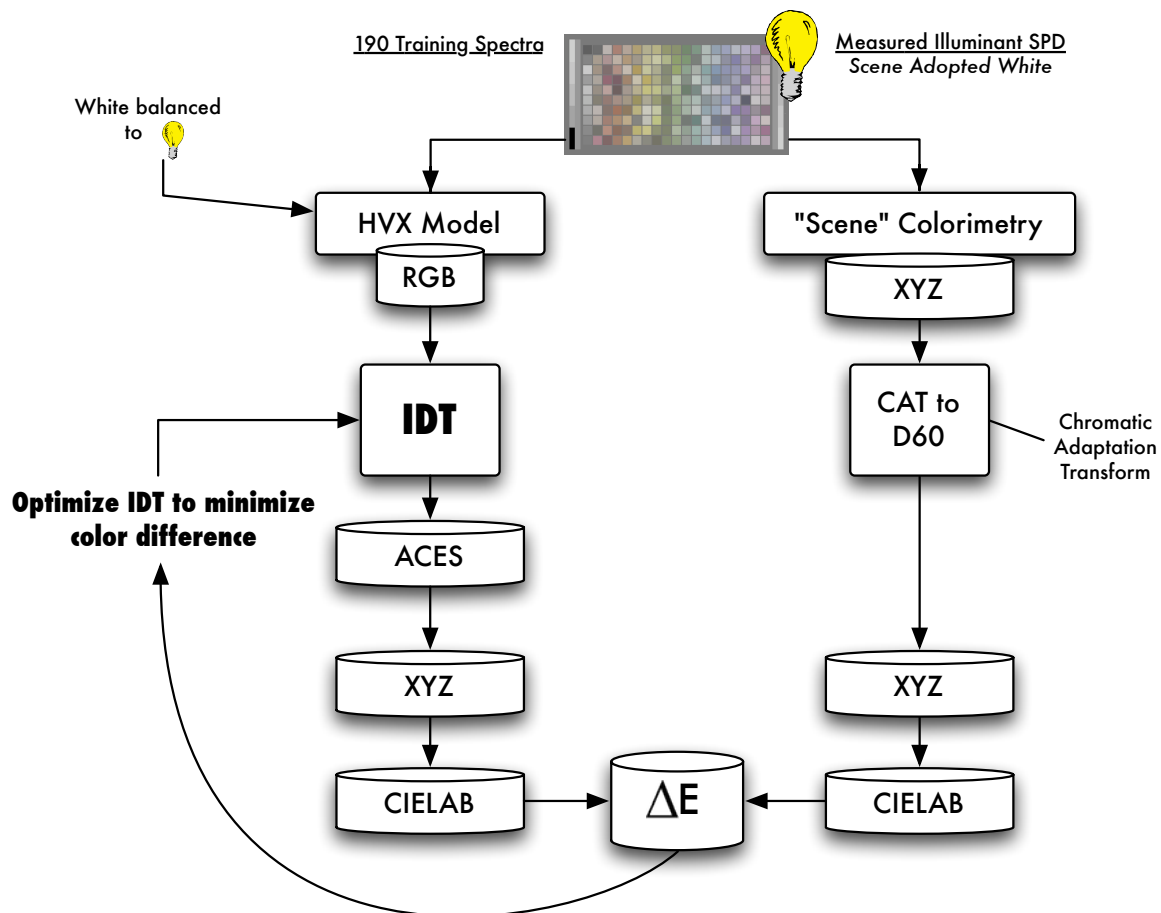


Fig. 16: IDT Method 1

Method #2: White Balance Only

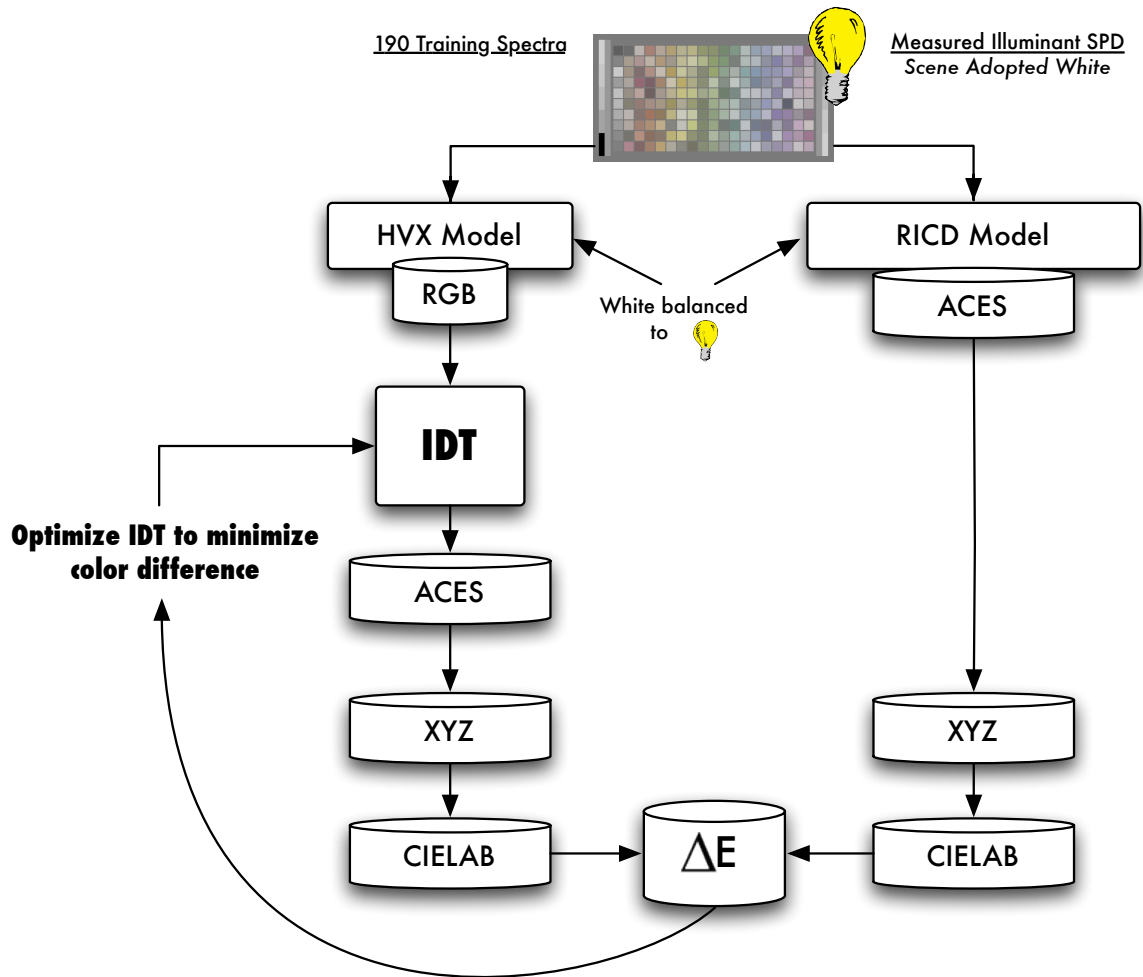


Fig. 17: IDT Method 2

4.2 Optimization

Functions were written in MATLAB for each stage of IDT computation.¹⁴ The following data was imported:

1. Spectral sensitivities for the Panasonic HVX, as derived in Section 3.2
2. Spectral sensitivities for the ACES RICD, as provided by Academy³
3. CIE 1931 color-matching functions¹³
4. 190 spectral reflectances, as training spectra

The data was imported and linearly interpolated to fit the HVX sensitivity data, which is based on wavelengths from 400 to 700 nm at 10 nm increments.

Optimization of the IDT was done in order to minimize the ΔE 2000¹³ between *a)* Camera white-balanced RGB multiplied by IDT to produce ACES and converted to XYZ *vs.* XYZ chromatically-adapted from scene adopted white to D60 for Method 1 and *b)* Camera white-balanced RGB multiplied by IDT to produce ACES values *vs.* ACES values captured with a RICD white-balanced to the scene adopted white for Method 2, for the training spectra.

ΔE 2000 (simply ΔE 00 hereafter) was chosen over ΔE because of its superior uniformity in quantifying color differences in terms of human visual perception. The MATLAB function `fmincon` was used with a trust-region-reflective algorithm to handle the regression. The function calculated the ΔE 00 for each of the training spectra and returned their mean as a single value. The optimization was, then, a minimization of the mean ΔE 00 from all training spectra, in this case the 190 patch reflectance set.

While a more extensive list of steps is laid out in a draft of the AMPAS IDT Documentation⁴, a brief description the functions written for the purposes of this paper provide an overview of the IDT computation and optimization process:

`Chromatic adaptation` — Computes the tristimulus values of the given scene illuminant to use as scene adopted white for generating a chromatic adaptation transform (CAT02 matrix⁴), which transforms the computed tristimulus values of the patches from the scene illuminant to a D60 illuminant. *Functions only used for Method 1.*

`Calculating white-balanced RGBs` — Computes two sets of white-balanced linear camera exposure values that would be produced if the camera and the RICD were to capture the patches under the scene illuminant given. *Function only used for Method 2.*

`Computing the average color difference`— Returns the mean ΔE 00 by computing and comparing the computed CIELAB values converted from `[rgbHVX * idtGuess]` and the chromatically-adapted tristimulus values of the patches. *Function used in both methods for use in regression loop.*

`Regression Loop` — Built-in MATLAB function to “Find minimum of constrained nonlinear multivariable function” where `cielab` was passed as the function `fun` of which the output should be minimized by modifying the `guess x0` values for the IDT.

4.3 Standard Illuminants as Scene Adopted White

As mentioned, a perfect blackbody illuminant was used for the first stage of IDT optimization. In the scenario of a manufacturer building an IDT for use with their camera, this would be the method of choice. They are building IDTs for industry-wide use, so they will likely use a standard illuminant such as a computed blackbody radiator at color temperatures that resemble those of typical lighting scenarios. IDTs were generated using 3200K and 5500K blackbody distributions (like those shown in Figure 15).

4.3.1 Results

The optimization was successful, and the 3200K and 5500K IDTs produced the IDTs shown below (average ΔE_{00} s of 1.83) for Method 1:

$$IDT_{3200K} = \begin{bmatrix} 0.7245 & 0.2918 & -0.0163 \\ 0.1111 & 0.9345 & -0.0456 \\ -0.0647 & 0.0346 & 1.0301 \end{bmatrix} \quad IDT_{5500K} = \begin{bmatrix} 0.6721 & 0.2892 & 0.0387 \\ 0.1038 & 0.9352 & -0.0390 \\ -0.0197 & 0.0177 & 1.0020 \end{bmatrix}$$

For Method 2 (with an average ΔE_{00} of 1.72):

$$IDT_{3200K} = \begin{bmatrix} 0.6768 & 0.1984 & 0.1247 \\ 0.1243 & 0.8476 & 0.0281 \\ -0.0218 & 0.0484 & 0.9734 \end{bmatrix} \quad IDT_{5500K} = \begin{bmatrix} 0.6459 & 0.2589 & 0.0953 \\ 0.1180 & 0.8856 & -0.0036 \\ -0.0072 & 0.0198 & 0.9874 \end{bmatrix}$$

4.4 Measured Illuminants as Scene Adopted White

The next step was to use a real illuminant as the scene adopted white for computing the IDT. Since one of the goals of this project is to provide feedback as to how well this type of camera can perform in a “best-case scenario,” the next step was to compute IDTs for real illuminants and see how well they perform in a real lighting situation. This process involved setting up typical lighting scenarios in Studio A at SOFA. The setup is described below. The spectral power distribution of an illuminant can be measured by aiming a spectroradiometer at a perfect diffuse reflector (PRD) lit by the illuminant. Using typical studio lights from the SOFA equipment cage, three illuminants were investigated for IDT computation and comparison.

4.4.1 Experimental Setup

The Gretag Macbeth Color Checker chart was set up next to a halon disc and several pieces of fruit as reference colors. This experimental setup is shown in Figure 18.

Three lights were used for the shoot: tungsten, HMI, and KinoFlo Daylight. The camera was white balanced by zooming the camera into the white halon disc (PRD). The spectroradiometer was then pointed at the PRD and spectral radiance was measured. The goal was to produce equal RGB code values for the PRD, since the spectral power distribution of the light will become the scene adopted white during optimization. Once the camera was white-balanced to the measured PRD, the scene was captured, making sure that no patches were clipping at high or low code values. The patches were then measured individually with the spectroradiometer to obtain spectral radiance. These steps were repeated for each lighting setup, making sure each light had proper time to warm up. The illuminant SPDs can be seen in Figure 19.

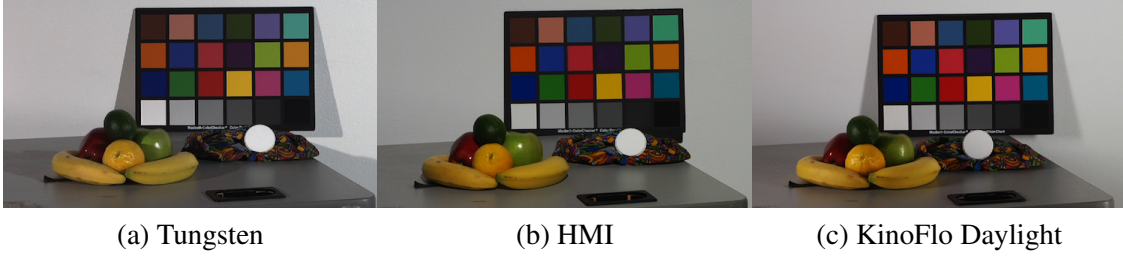


Fig. 18: The footage from the experimental setup for IDT computation. The scene contains a Gretag Macbeth ColorChecker, a halon disc acting as a perfect diffuse reflector, and some fruit as reference colors.

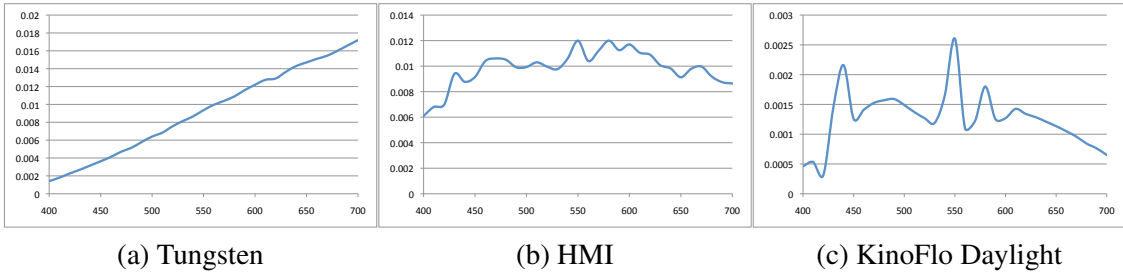


Fig. 19: The spectral power distributions for Tungsten, HMI, and KinoFlo Daylight, as measured off of the PRD disc with a spectroradiometer.

4.4.2 Results

The measurement for the PRD could now become the scene adopted white by using its spectral power distribution as the scene illuminant in the IDT optimization functions. Using these new scene adopted whites from measurements of studio-tungsten, HMI, and KinoFlo Daylight, the following IDTs were computed for Method 1 (average ΔE_{00} of 1.68):

$$IDT_{Tung} = \begin{bmatrix} 0.6722 & 0.2102 & 0.1176 \\ 0.0989 & 0.8779 & 0.0232 \\ -0.0512 & 0.0547 & 0.9965 \end{bmatrix} \quad IDT_{HMI} = \begin{bmatrix} 0.6604 & 0.2614 & 0.0782 \\ 0.1143 & 0.9009 & -0.0152 \\ -0.0158 & 0.0188 & 0.9970 \end{bmatrix}$$

$$IDT_{KinoDay} = \begin{bmatrix} 0.6279 & 0.2947 & 0.0774 \\ 0.1039 & 0.9084 & -0.0124 \\ -0.0111 & 0.0116 & 0.9995 \end{bmatrix}$$

For Method 2 (average ΔE_{00} of 1.75):

$$IDT_{Tung} = \begin{bmatrix} 0.7280 & 0.2933 & -0.0213 \\ 0.1113 & 0.9349 & -0.0462 \\ -0.0750 & 0.0390 & 1.0359 \end{bmatrix} \quad IDT_{HMI} = \begin{bmatrix} 0.6765 & 0.2906 & 0.0330 \\ 0.1137 & 0.9224 & -0.0361 \\ -0.0209 & 0.0167 & 1.0042 \end{bmatrix}$$

$$IDT_{KinoDay} = \begin{bmatrix} 0.6378 & 0.2989 & 0.0633 \\ 0.1002 & 0.9225 & -0.0227 \\ -0.0141 & 0.0105 & 1.0036 \end{bmatrix}$$

After ingesting the video footage, the code values were extracted for the 24 patches and the PRD for each of the three lighting setups. A new function `idt_check` was written for this next stage. These gamma-corrected camera code values were linearized using the $\gamma = 0.72$ power function, and run through the IDT that had been generated previously. Those values come out of the IDT as ACES values, which are then converted to CIELAB. The function also uses the spectral measurements for the patches and PRD to calculate actual CIELAB values for the 24 patches. The function returns ΔE_{00} differences for each of the 24 patches. The diagram in Figure 20 gives a general overview of what the function computes (diagram refers to Method 1: Chromatic Adaptation).

It must be noted that linearizing the HVX code values and converting to ACES RGB relative exposures creates values that are essentially unit-less. The Academy's recommendation for this is to balance the camera so that a neutral 18% grey card will created ACES RGB = 0.18. Since there was no gray card in the scene shot, the patch that GretagMacbeth defines as a 20% reflector (Patch #22) was used to normalize the ACES values to 0.20.

The average ΔE_{00} value for the 24 color patches was 2.31, and ranged from 0.48 to 7.7 for the Method 1: IDT Tungsten. The results for Method 2: IDT Tungsten were similar with an average ΔE_{00} of 2.67. The most concerning problem, however, can be seen in patches 19-24 in Table 2. Patch 22 has a low ΔE_{00} as expected, due to the normalization. However, as the grayscale goes up or down, the ΔE_{00} quickly rises. Errors in bright colors can be expected because of the inherent colorimetric differences between cameras, but these errors, seen in the neutral tonescale, are unacceptable. Given the distribution of errors around normalization, it can be inferred that the gamma estimation of 0.72 was incorrect.

4.5 Gamma Fitting

The errors seen in the neutral patches were an indication that a more accurate gamma estimation exists. It was no surprise, given the errors seen in the data gather during grayscale characterization (see Section 3.1). There must exist a power function that better fits the actual tonescale reproduction of the camera. Therefore, a new optimization function was created.

The function even further minimizes ΔE_{00} based on trying different gamma values for linearization. Because the spectral sensitivity curves derived for the HVX assumed the 0.72 power function, this became a game of trial and error. Changing the gamma for the sensitivity curves created a lower ΔE_{00} during IDT computation. Using that newly computed IDT, a gamma optimization lowered the ΔE_{00} even further and returned a new gamma. This process was repeated until the IDT computation produced the lowest possible average ΔE_{00} without starting to increase again. The gamma that produced this value was $\gamma = 0.64$.

The new gamma and IDT was re-analyzed using the ΔE_{00} from the experimental setup. Sure enough, the grayscale lined up almost perfectly.

After re-optimizing and settling on a new gamma value, the IDTs must be recalculated. They are as follows:

Method 1: Chromatic Adaptation

Experimental Setup: Real Illuminant

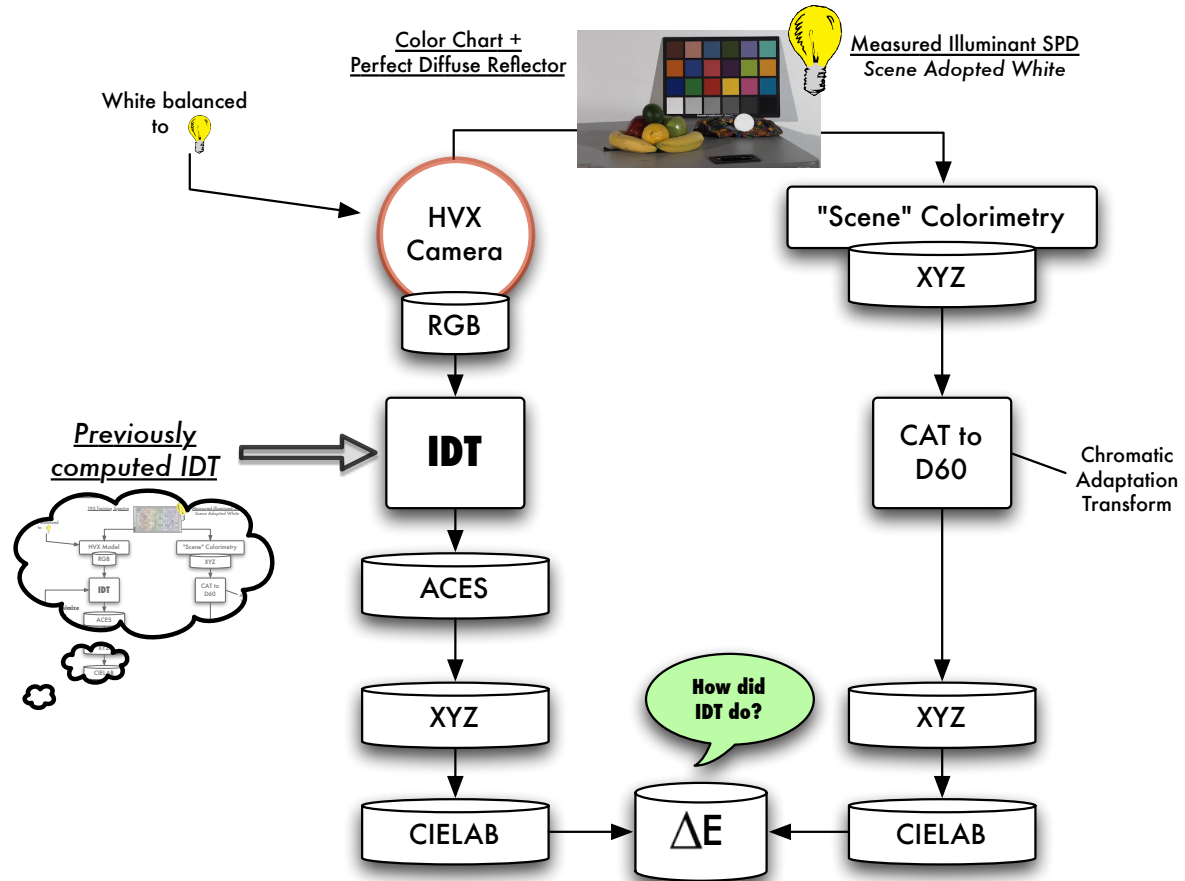


Fig. 20: The method for seeing how well the previously computed IDT performed under a real lighting situation. Note that the HVX Camera is the real video code values coming out of the camera.

Table 2: This table represents the results from the “IDT check” stage(with gamma 0.72) with Method 1 IDT for Tungsten. In bold are the grayscale values that indicated an improper gamma value.

Patch	Measured			HVX			
#	L*	a*	b*	L*	a*	b*	ΔE_{00}
1	36.86	14.44	14.63	38.06	15.67	16.46	1.55
2	64.37	18.79	15.96	64.54	17.42	15.99	0.97
3	48.26	-4.01	-21.12	47.04	-4.46	-22.43	1.42
4	40.56	-16.04	17.41	41.15	-15.63	19.95	1.63
5	56.13	11.9	-23.8	55.88	12.84	-20.84	2.45
6	69.88	-32.58	-5	67.87	-29.68	-3.86	2.08
7	61.08	30.89	54.53	60.28	29.16	56.62	1.76
8	38.8	11.73	-38.97	37.10	15.57	-41.37	2.27
9	52.51	48.92	18.45	54.06	44.79	20.29	2.63
10	31.4	23.13	-17.81	32.52	24.24	-15.48	1.91
11	71.2	-27.29	47.61	69.88	-25.34	49.90	1.83
12	73.78	15.82	60.68	71.47	12.68	61.29	2.70
13	28.57	16.34	-43.91	25.20	22.78	-51.41	3.45
14	52.38	-43.82	25.55	53.04	-45.13	26.47	0.81
15	44.49	58.24	28.33	49.09	50.91	36.81	7.71
16	82.39	1.65	68.92	79.55	-0.37	71.02	2.36
17	54.45	53.52	-8.89	56.68	47.89	-4.61	3.27
18	49.7	-22.19	-31.5	49.41	-21.85	-30.51	0.48
19	92.56	0.04	1.89	86.61	0.49	0.58	4.00
20	80.23	-0.15	-0.06	76.38	0.12	-0.37	2.76
21	66.82	-0.38	-0.22	64.53	0.16	-0.44	2.04
22	52.11	-0.26	-0.45	51.83	0.19	-0.55	0.72
23	36.4	-0.4	-0.44	35.46	0.56	-1.50	1.91
24	22.07	-0.08	-0.68	19.35	0.82	-2.23	2.71

$$IDT_{Tung} = \begin{bmatrix} 0.6775 & 0.1964 & 0.1262 \\ 0.1239 & 0.8449 & 0.0312 \\ -0.0252 & 0.0553 & 0.9700 \end{bmatrix} \quad IDT_{HMI} = \begin{bmatrix} 0.6482 & 0.2493 & 0.1025 \\ 0.1270 & 0.8694 & 0.0035 \\ -0.0068 & 0.0199 & 0.9869 \end{bmatrix}$$

$$IDT_{KinoDay} = \begin{bmatrix} 0.6130 & 0.2846 & 0.1024 \\ 0.1134 & 0.8809 & 0.0057 \\ -0.0047 & 0.0136 & 0.9911 \end{bmatrix}$$

Method 2: White Balance Only

$$IDT_{Tung} = \begin{bmatrix} 0.7328 & 0.2698 & -0.0026 \\ 0.1456 & 0.8858 & -0.0314 \\ -0.0373 & 0.0410 & 0.9963 \end{bmatrix} \quad IDT_{HMI} = \begin{bmatrix} 0.6588 & 0.2727 & 0.0685 \\ 0.1316 & 0.8774 & -0.0090 \\ -0.0080 & 0.0186 & 0.9894 \end{bmatrix}$$

$$IDT_{KinoDay} = \begin{bmatrix} 0.6162 & 0.2851 & 0.0987 \\ 0.1137 & 0.8834 & 0.0029 \\ -0.0048 & 0.0134 & 0.9915 \end{bmatrix}$$

4.6 Testing the IDTs

The Input Device Transforms shown above represent the best working versions for each IDT type. For meaningful analyses of IDTs, it is necessary to know the computational process that was used to build it. They were built with three real, measured illuminants. These will herein be referred to as *design illuminants*, and this is indeed similar to a film stock have a design illuminant of tungsten or daylight. In addition to the illuminants, the training spectra consisted of 190 training spectra, which representative and distributed similar to real-world surfaces that exist in images. In other words, the IDT will perform best if those 190 training patches happened to be in front of the camera under one of the design illuminants.

The ΔE_{00} values will vary depending on the degree at which the scene resembles the IDT design process. Since the goal was to determine what kind of results could be achieved using the IDTs, seven conditions for analysis of color difference were investigated:

Condition A: *Training Spectra + Design Illuminant* – This is the scenario that was used in the beginning to minimize ΔE_{00} and compute an IDT. This can be thought as a nearly impossible “best-case scenario” to test to see how well it performs in its optimal state.

Condition B: *Different Spectra + Design Illuminant* – In the case of this paper, the different spectra of Condition B consisted of the spectral radiance measurements made during the Experimental Setup. Camera exposures are computed using the HVX Model, run through the IDT, and compared to the ACES values as measured off of the spectra. Note that the camera exposures are *computed* instead of actually shooting the spectra. This avoids camera variability, but spectral measurement error is still a factor.

Condition C: *Different Spectra + Design Illuminant + Real World Variability* – The camera and lighting equipment on-set can vary immensely from day-to-day. In the case of this research, this condition was the experimental setup. The actual camera video signal was run through the IDT, and compared with the ACES values that come from spectral readings. This is

similar, but not quite as rudimentary as some productions might go. The design illuminants are still being used, and the setup was done in a somewhat controlled, studio environment. Variability is still a major factor for this condition.

4.6.1 Results

The results of analyzing these three conditions are shown in the the table below (Condition A), Table 4 (Condition B), and Table 5 (Condition C). As expected, the average ΔE_{00} increased with the variability from the design of the IDT.

Table 3: Below are the regression results from Condition A. During computation, this was the minimum average ΔE_{00} at the end of regression. The values represent the minimum ΔE_{00} that the regression algorithm was able to achieve with the given IDT computation.

Average ΔE_{00} from Condition A			
	Tungsten	HMI	KinoFlo
Method 1	1.8	1.7	1.6
Method 2	1.9	1.7	1.6

5 DISCUSSION

The results of the initial IDT computation with an average ΔE_{00} of 1.83 was encouraging. There was evidence that the optimization was indeed working. The difference between Method 1 and Method 2’s IDTs was as expected. They are going to render colors differently. In addition, given the difference between the HVX spectral sensitivities and CMFs, the color differences were expected.

The inner workings of the camera likely included matrices and look-up tables that were hard to predict. Once the grayscale ΔE_{00} s were fixed with the gamma-fitting, it can be speculated from the CIELAB plots on what causes these color differences. Given the systematic differences in the CIELAB plots, one cause could be due to internal processing that is happening as part of the spectral sensitivity curves. If the gamma is affecting the spectral sensitivity curves inside the camera, then the hopes of a versatile IDT for the HVX are slim. Since there is inherently a matrix happening inside the camera, there is the question of where it is applied. If it is applied after the gamma, the IDT will have much worse results than after the gamma, since the IDT will be compensating for non-linear changes in a linear space.

Although the same settings were used for each set-up, the ΔE_{00} differences for Condition C (Experimental Setup) were very high for HMI and KinoFlo lights. The results were troubling, since the IDTs were designed for those lights. Referring to their SPDs, the “spikey” nature could have to do with this difference. More trials are needed to investigate this.

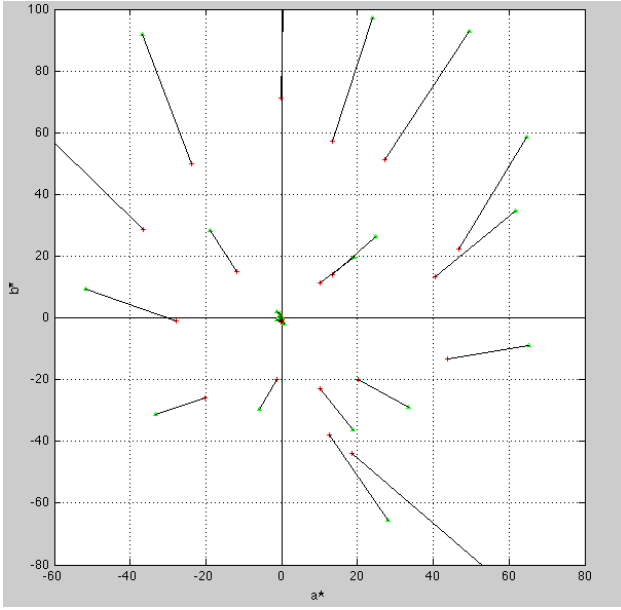
The results show that the overall practicality of this set of IDTs is limited to a small degree of flexibility. While it works great in standard illuminants and hypothetical scenarios, the HVX and its “blackbox” of processing cause trouble when actually shooting. Camera companies are striving to create a camera that is reliable and versatile, while overcoming the competitors. To achieve this, it might be beneficial for them (and consumers) to have on-the-fly processing going on to clip

Table 4: Below are the ΔE_{00} results for Condition B for the 24 colors patches.

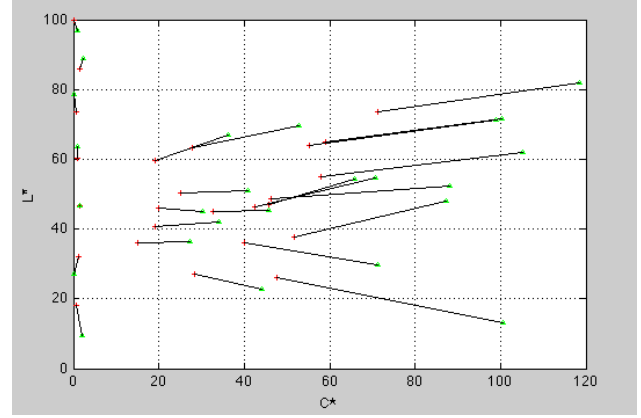
Method 1: Chromatic Adaptation				Method 2: White Balance Only			
Patch	Tungsten	HMI	KinoFlo	Patch	Tungsten	HMI	KinoFlo
1	0.4	0.6	0.7	1	1.8	0.4	0.3
2	2.5	1	1.3	2	2.6	1.1	0.5
3	1	0.5	0.9	3	4	1.1	0.4
4	1.5	1.1	0.9	4	1.4	1	1.3
5	1.2	1.1	1.4	5	2.2	0.7	1.1
6	3.5	1.6	1.5	6	0.8	1.2	2.1
7	4.3	4	4.1	7	2.9	3.1	3.8
8	0.7	0.3	0.5	8	6.1	1.5	0.3
9	1.6	1.8	2.3	9	4.7	2.2	1.3
10	2.1	1.5	2	10	2.2	2	2.2
11	1.6	0.9	0.6	11	2	0.8	0.7
12	3.3	2.7	3.2	12	3.4	2.2	2.4
13	1	1.2	1.3	13	7.2	2.2	1.2
14	2.3	1.4	0.9	14	1.3	1.3	1.5
15	5.2	4.4	5.1	15	8.1	4.7	4.1
16	1.6	0.2	0.3	16	3.2	0.5	0.8
17	2.2	2.8	3.3	17	3.8	2.8	2.5
18	4.5	2.3	2.5	18	1.7	1.7	3.3
19	0.2	0.4	0.1	19	0.8	0.6	0.7
20	0.1	0.3	0.2	20	0.3	0.3	0.3
21	0.2	0.2	0.2	21	0.2	0.2	0.1
22	0.2	0.1	0.1	22	0.2	0.2	0.1
23	0.2	0.1	0.2	23	0.2	0.1	0.1
24	0	0	0.1	24	0.2	0.1	0.1
Average	1.7	1.3	1.4	Average	2.5	1.3	1.3

Table 5: Below are the ΔE_{00} results for Condition C for the 24 colors patches.

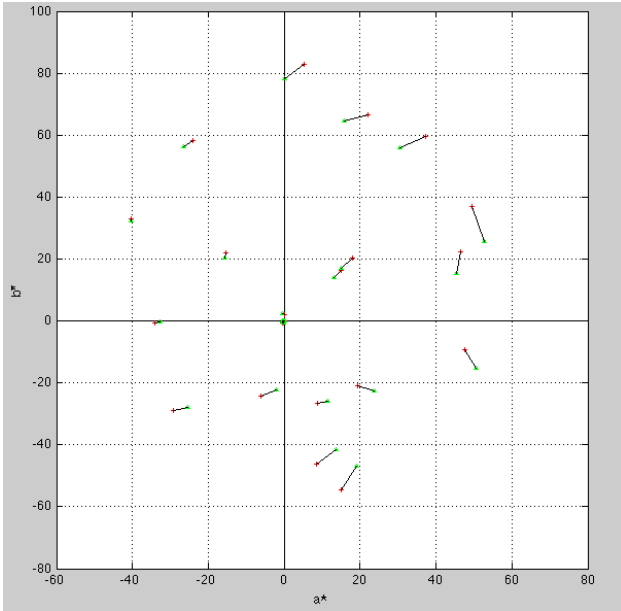
Method 1: Chromatic Adaptation				Method 2: White Balance Only			
Patch	Tungsten	HMI	KinoFlo	Patch	Tungsten	HMI	KinoFlo
1	2.3	6.7	10.1	1	5.3	3.8	4.9
2	2.2	9.7	11.5	2	3.5	1.1	1.7
3	2.6	6.8	7.2	3	4.3	4.6	1.9
4	2.3	7.1	8.2	4	4.9	3.9	6.5
5	2.1	5.8	5.8	5	5.6	1.1	3
6	0.9	11.3	8.7	6	8.9	4.3	5.8
7	2.7	11.8	14.8	7	4.5	4.1	4.8
8	3.8	8.9	8.8	8	2.7	3.8	1.4
9	3.8	13.1	16.5	9	5.3	4.2	2.3
10	2	7	6.7	10	6.9	2.4	3
11	2.8	11.5	8.8	11	7.6	4.7	5.4
12	2.5	10.5	10.4	12	6.5	4.4	7.5
13	5.9	14.8	13.1	13	6.2	5	4
14	2.4	11.3	10.7	14	4.9	5.1	5.7
15	9.3	16.7	24.4	15	9.4	8.4	7.8
16	2.9	10.7	8.9	16	8.4	5	6.5
17	3.9	10.1	13.1	17	7	3.2	1.9
18	1	5.8	7.2	18	3.1	4.6	3.3
19	1.5	2.3	3.1	19	11.3	10.2	10.4
20	0.6	3.7	4	20	8.2	6.1	6.9
21	1	3.2	4.4	21	4.8	3.1	3.5
22	0.8	2.3	2	22	0.4	1.7	1.6
23	2.8	4	4	23	3.1	1.6	2.1
24	4.3	5.7	6.4	24	3.6	1.6	1.8
Average	2.8	8.4	9.1	Average	5.7	4.1	4.3



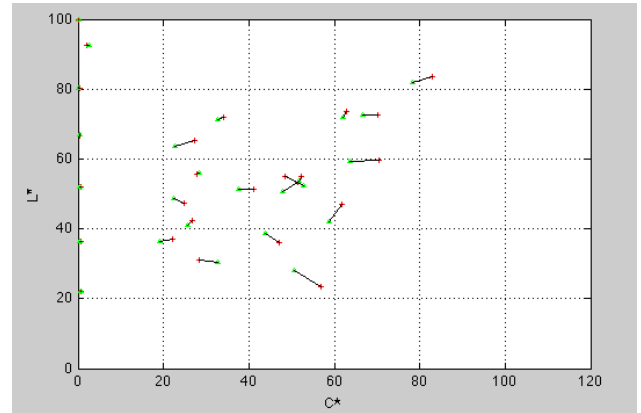
(a) Method 1 + Condition C + HMI IDT



(b) Method 1 + Condition C + HMI IDT



(c) Method 2 + Condition B + Tungsten IDT



(d) Method 2 + Condition B + Tungsten IDT

Fig. 21: Figure 21a was generated using the CIELAB values as shown earlier in Figure 20 for the Method and Condition shown. a^* vs. b^* on the left and L^* vs. C^* on the right (Method 1 = Chromatic Adaptation, Method 2 = White Balance Only, Condition B = Different spectra, Design illuminant, Condition C = Real-shooting scenario with design illuminant and spectra from Condition B)

highlights, not oversaturate the colors, or other uncontrollable circumstances that the camera can deal with.

6 CONCLUSION AND THOUGHTS

The Academy has been struggling to finalize an IDT document for these reasons. The circumstances of shooting are hard to predict. In addition to variability, what happens when someone shoot under tungsten because they want to have a “warm” scene? If a tungsten IDT were used, it would be rendered neutral. That is not to say that the IDT system will not work. It is similar to film in that you will have certain balances to choose from. Once a decision has been made on IDT design, their implementation will soon follow. Deciding where an exactly in the workflow an IDT should be applied (in or out of camera) is not a decision that the Academy should be concerned with. The industry is going to use them in whatever way works best for them.

For the School of Film and Animation at RIT, it will be a long road to a calibrated and properly color-managed workflow. The equipment will not accommodate this kind of professional framework at this point, which is acceptable. Currently, a Rec. 709-calibrated workflow would be more possible. The results of this paper provide proof of a problem that arises from inside these kinds of cameras. Even with extensive characterization, there was only a certain level of control possible before the camera acts on its own.

As the Academy finishes up their work on an IDT document, there is a list of considerations to be made. Fitting “prosumer” cameras into the framework is not likely at the top of that list. It is, however, important not to leave anyone out on such as important standard. Once an IDT design and implementation method is decided upon, the world of digital color management will be much easier.

References

- [1] D. L. Long, “Personal communication,” 2008-2010.
- [2] Science and A. Technology Council, “Image interchange framework,” 2010.
- [3] A. Forsythe, J. Goldstone, E. Giorgianni, J. Houston, R. Feeney, and L. Borg, “Academy color encoding specification (aces).” 2008.
- [4] J. Goldstone, *Recommended Procedures for the Creation and Use of Digital Camera Input Device Transforms (IDTs)*. Los Angeles: Science and Technology Council, Academy of Motion Picture Arts and Sciences (A.M.P.A.S.), 2009.
- [5] E. J. Giorgianni and T. E. Madden, *Digital color management : encoding solutions*. Hoboken, NJ: J. Wiley, 2nd ed., 2008. 2008041558 by Edward J. Giorgianni & Thomas E. Madden. Includes bibliographical references and index.
- [6] Panasonic, “Panasonic hvx-200 user manual,” 2005.
- [7] C. A. Poynton, *Digital video and HDTV : algorithms and interfaces*. Morgan Kaufmann series in computer graphics and geometric modeling, Amsterdam ; Boston: Morgan Kaufmann Publishers, 2003. 2002115312 Charles Poynton. ill. ; 24 cm. Includes bibliographical references and index.

- [8] M. D. Fairchild, *Color appearance models*. Wiley-IS&T series in imaging science and technology, Chichester, West Sussex, England ; Hoboken, NJ: J. Wiley, 2nd ed., 2005. 2004043111 Mark D. Fairchild. ill. (some col.) ; 25 cm. Includes bibliographical references (p. [361]-377) and index.
- [9] H. H. Hugh S. Fairman, Michael H. Brill, "How the cie 1931 color-matching functions were derived from wright-guild data," *Color Research and Application*, vol. 22, no. 1, 1996.
- [10] W. contributors, "Cie 1931 color space," 2010.
- [11] R. Patterson, "Evaluating density metrics for scanning motion picture negatives," *SMPTE Motion Imaging Journal*, vol. May/June Issue, 2008.
- [12] D. L. Long, "Digital cinema course notes," 2008-2009.
- [13] M. C. S. Laboratory, "Useful color data."
- [14] S. Westland and C. Ripamonti, *Computational colour science using MATLAB*. Hoboken, NJ: J. Wiley, 2004. 2003021483 Stephen Westland, Caterina Ripamonti. ill. ; 25 cm. Includes bibliographical references (p. [199]-203) and index.