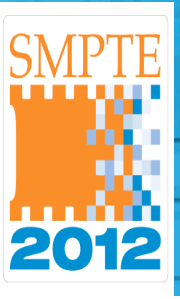
The Unfolding Merger of Movie and Television Technology

Gary Demos
Image Essence LLC



Disappearing Anisotropy in Television Technology

- Interlace is waning (at least it is hoped that it is…!!!)
- Scanlines are no longer prominent
- 4:2:2 is giving way to 4:4:4 (although 4:2:0 remains)
- RGB and XYZ in DCinema have full resolution and are more equal in coded bits, with some trends in television moving in that direction
- Displays and projectors usually offer 2D uniform rasters (without scanlines)
- Cameras have offered 3-Chip R, G, and B, and several single-chip cameras exist with higher uniformity than the Bayer pattern
- Frame-shuttered cameras are available which do not scan with a "rolling shutter"

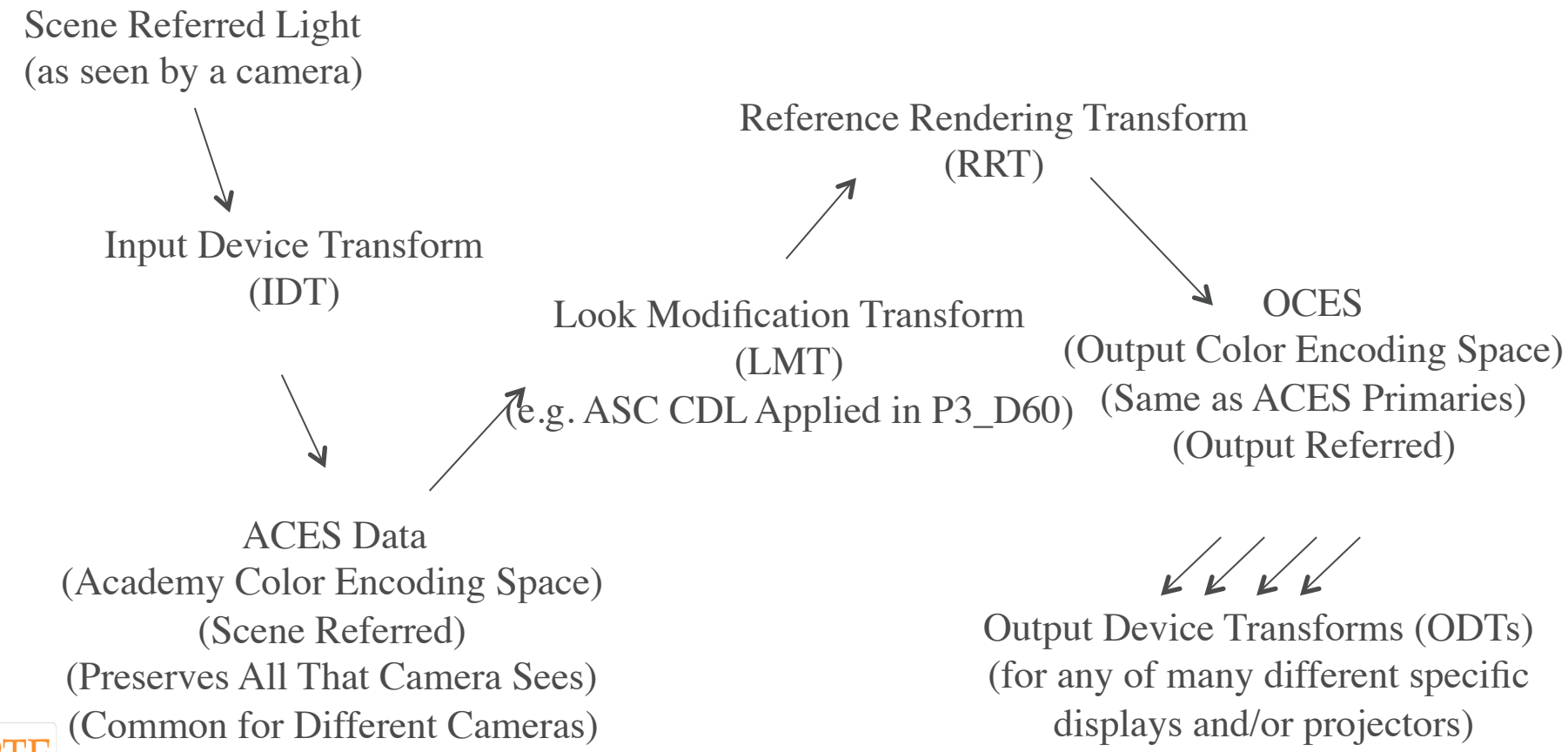
Frame Rates

- Film and DCinema have been 24fps
- 25fps on 50Hz television systems
- 3-2 pulldown is problematic (to un-do or transcode if mixed cadence or composited with mixed elements)
- Sports requires a higher frame rate
- 3D has issues when using 24fps
- DCinema is looking at 48fps, 60fps, 96fps, 120fps, and other rates, with 48fps and 60fps being considered strongly for 3D
- The exploration of higher frame rates in DCinema is implicitly beginning to merge a key system attribute with television

Rec709

- In-camera rendering to create output-referred images
- In-camera "painting" of exposure, color, etc.
- Typical in-camera rendering ingredients include matrix, gamma boost, "knee" for highlights, and often some black adjustment
- Rec709 signal from the camera is ready to plug into the Rec709 input of a display or projector
- Gamma 2.22 (1/.045) with linear black toe below about 1% is fairly meaningless
- Recently a Gamma 2.4 was specified for display presentation

ACES System


Scene Referred Light
(as seen by a camera)
Input Device Transform
(IDT)
ACES Data
(Academy Color Encoding Space)
(Scene Referred)
(Preserves All That Camera Sees)
(Common for Different Cameras)
Look Modification Transform
(LMT)
(e.g. ASC CDL Applied in P3_D60)
Reference Rendering Transform
(RRT)
OCES
(Output Color Encoding Space)
(Same as ACES Primaries)
(Output Referred)
Output Device Transforms (ODTs)
(for any of many different specific
displays and/or projectors)

Recommendation for Applying ASC_CDL to ACES

ACES → Matrix to P3_D60 → ASC_CDL → Matrix (inverse) back to ACES

ASC_CDL 1.2:

1) r,g,b gain (1.0, 1.0, 1.0 is no adjustment)
2) r,g,b offset (0.0, 0.0, 0.0 is no adjustment)
3) r,g,b gamma (1.0, 1.0, 1.0 is no adjustment)
4) saturation adjust (0.0 is black and white, 1.0 is no adjustment, >1.0 is a saturation boost)

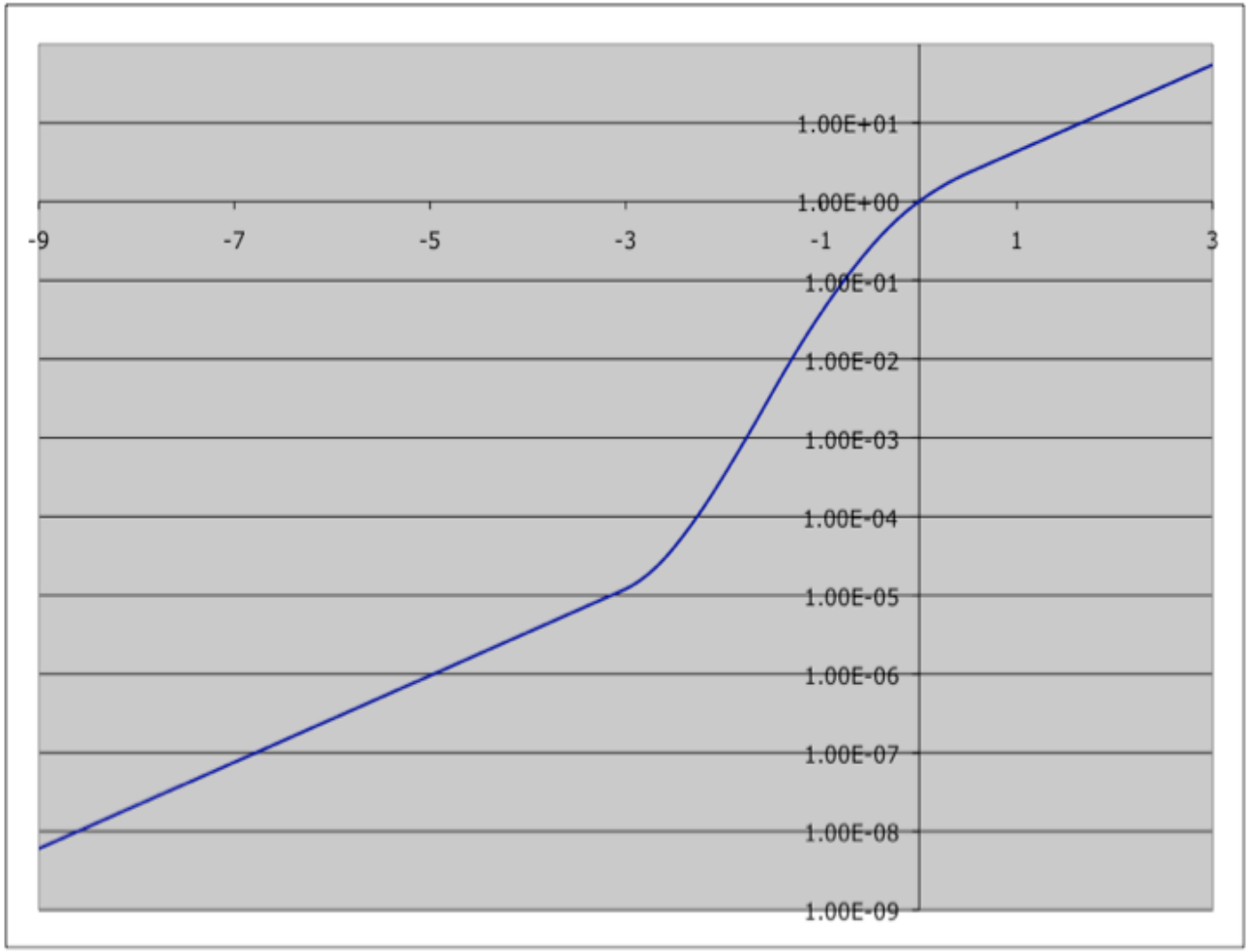
RRT Development

- Initial RRT was based upon a film density model
- Extended beyond film's range of color and brightness combinations (maximum red, green, and blue on film print are two layer combinations, and thus are inherently darker than red, green, and blue on an additive display or projector)
- Begin exploration of high dynamic range (both darker and brighter than projected film's range)
- A new matrix + tone curve additive (non-density) empirical model, with red, yellow, and blue hue region contrast adjustments
- Explorations of ratio restoration using norms with the tone curve(s)
- Replaced the hue region contrast adjustments with curves, using a cosine blended overlap as a function of hue angle

RRT/ODT Tools and Testing Example Configuration

- Clip ACES up to zero
- Tone curve with ratio restore (R, G, B ratios restored vs norm with ~1.5 gamma boost)
- Dynamic matrix, with off-diagonal term reduction driven by hue region and saturation, using slow_In_out
- Hue region brightness curves for red, yellow, and blue
- Hue region lifting of brightness and saturation boost for red and yellow face shadow regions
- Asymptotic high brightness desaturation for desaturated colors, using slow_In_Out
- RRT output in the Output Color Encoding Space (OCES) in ACES primaries
- ODTs with tone curve and partial ratio restore for highlights having saturated colors

RRT Tone Curve

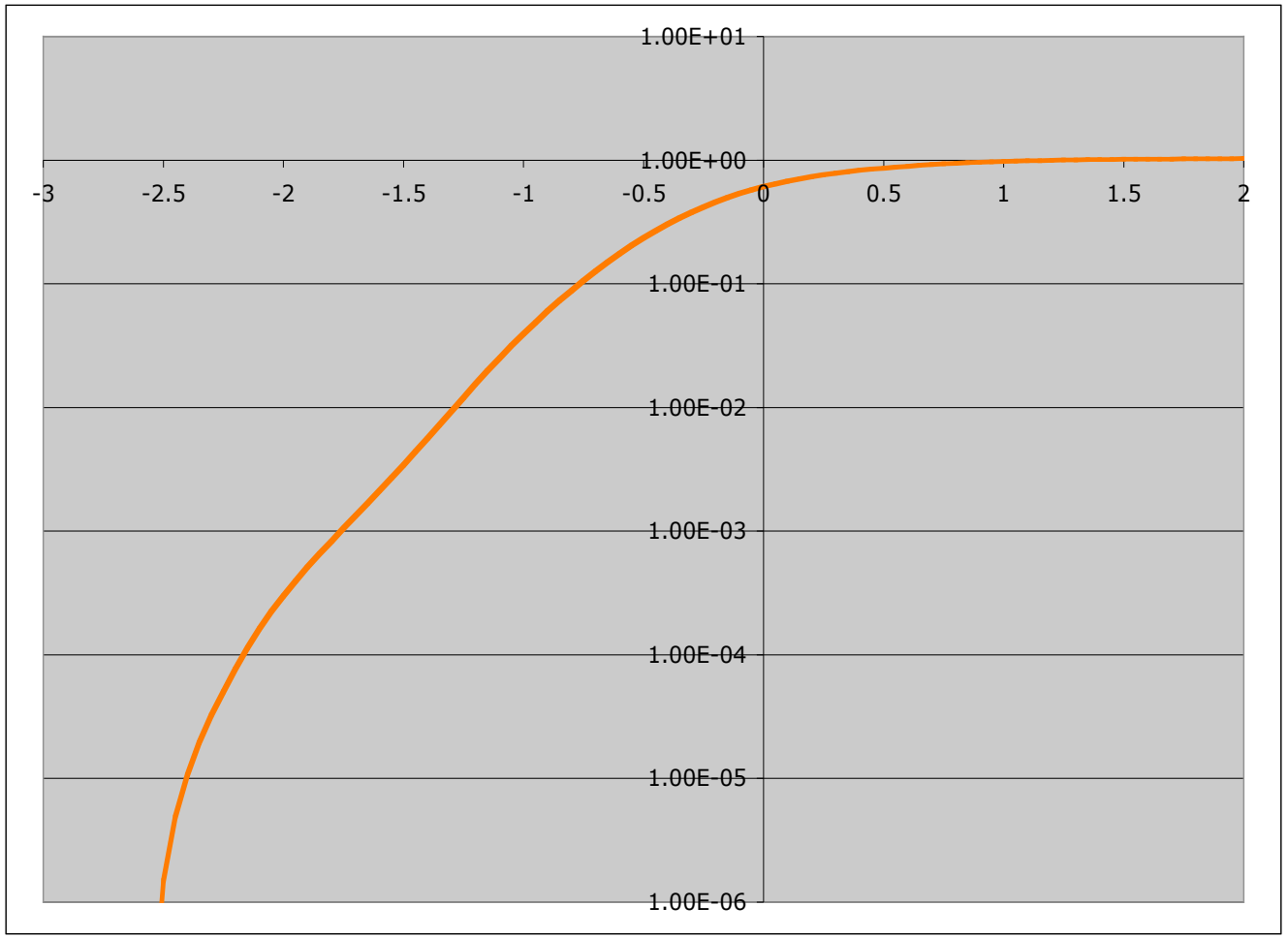


Independent R, G, B lookup, or norm lookup

For norm lookup, useful norm is $(r^2+g^2+b^2)/(r+g+b)$

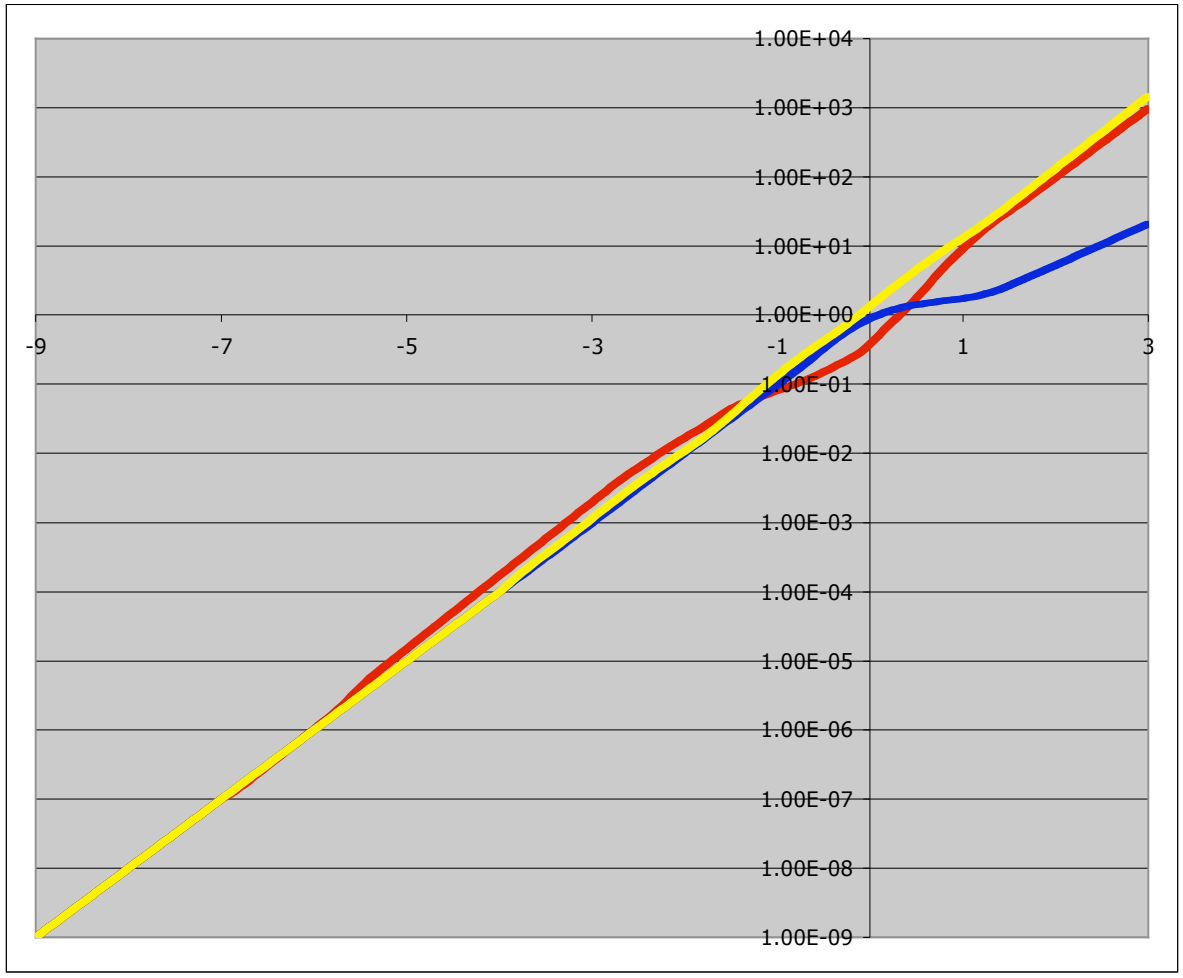
Also ~1.5 gamma boost of r, g, and b with respect to the norm lookup

ODT Tone Curve (non-HDR)



Note:
This looks like an S-Curve in a linear vs linear plot

Red, Yellow, and Blue Hue Region Curves



RRT Matrices

- Example static matrix:

```
Float MTX[9] = { \
 0.926,  .038,  .036; \
 .130,  .800,  .070; \
-.035, -.035, 1.070};
```

row ordering:
*R_out = .962*R_in+.038*G_in+.036*B_in*
*G_out = .130*R_in+.800*G_in+.070*B_in*
*B_out = -.035*R_in-.035*G_in+1.070*B_in*

Note: all positive off-diagonal matrix terms yield desaturation and limit maximum saturation

- Example dynamic matrix:

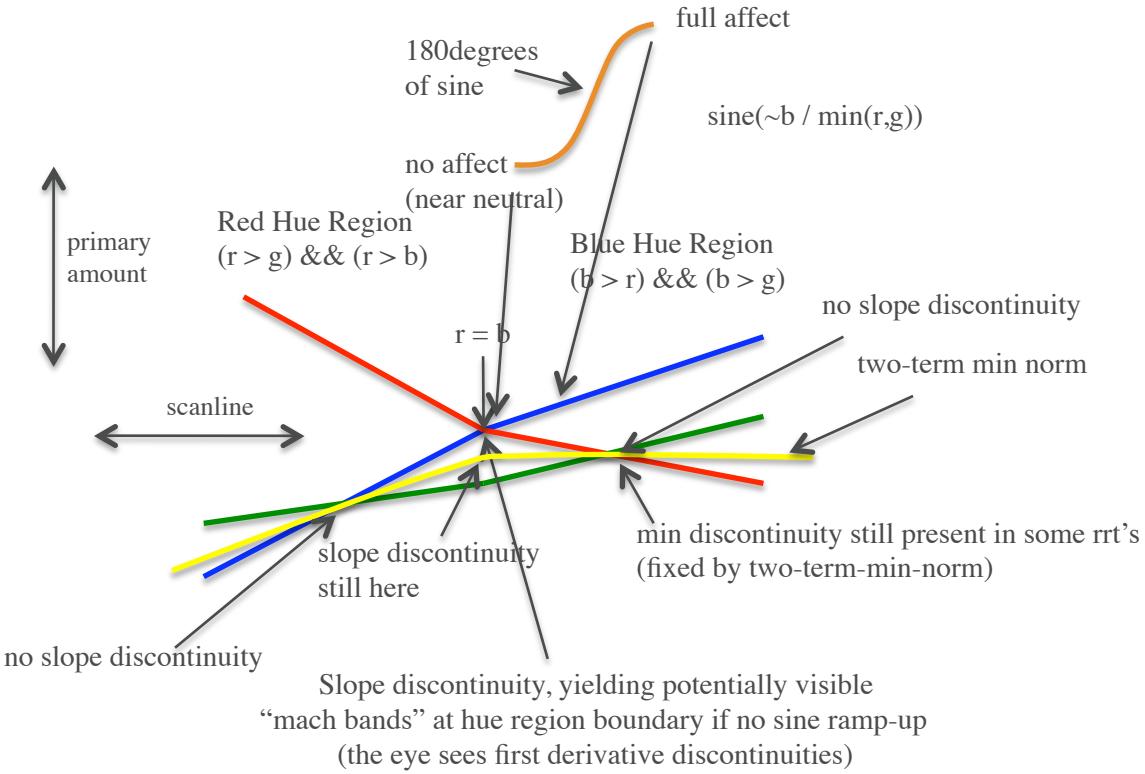
```
float MTX[9] = { \
 0.906, 0.058, 0.036, \
 0.1467, .8133, .04, \
-.06, -.037, 1.097 };
```

red in green term:
when larger makes reds more orange,
when smaller makes reds less orange

The blue row negative terms boost saturation of red, orange, yellow, and green and slightly alter their hue

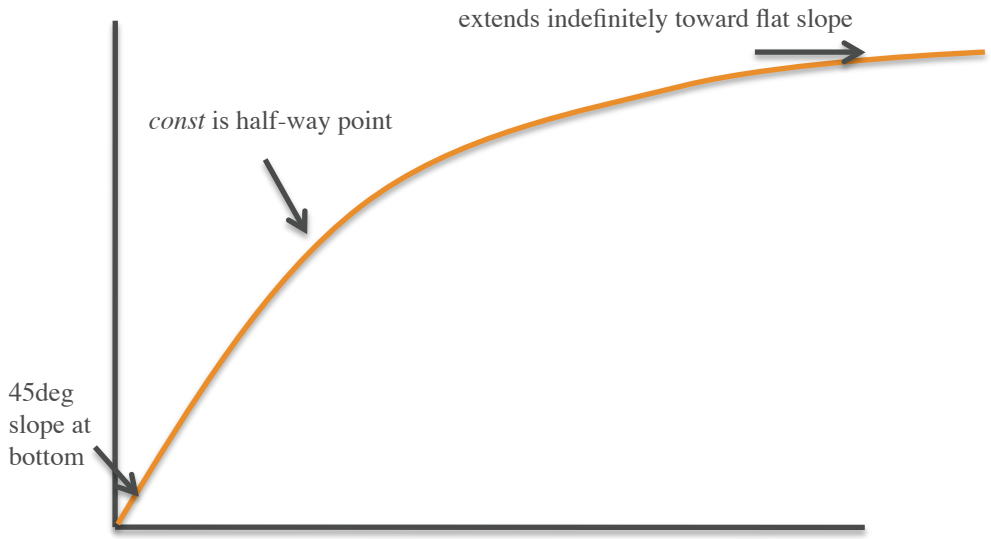
Note: dynamic matrix off-diagonal terms have reduced magnitude according to hue region by ramping these out with a smooth slow-in-out function going away from neutral.

Reasons for sine/cosine function ("slow in out") and two-term-min-norm



notes: although neutral is smooth, potential for discontinuities within hue region.
The max is solved in hue regions using "slow in out" sine function across neutral.
The middle and min crossover issue in hue-regions is solved by two-term-min-norm.
Slope discontinuity still present for min-norm at max crossover (thus when switching between hue regions).

Asymptotic function (used for saturation adjustment):



Formula: x / (*const* + x)

only one adjustment: *const*

(can further shape with pow)

ACES negative values

- Small negative values inherently valid for data for "capped lens" black when a camera is calibrated to *average* 0.0. In order to *average* 0.0, it is necessary to go slightly negative as well as slightly positive, due to the presence of thermal noise in the sensor.
- Large negative values are possible for some saturated colors due to oversimplified Input Device Transform (IDT) matrix processing, and due to camera spectral sensing functions not being linearly transformable into CIE 1931
- Pushing down ("crushing") blacks as an adjustment can also yield significant large regions of negative values

Negative Values in the RRT/ODT

- One experimental rrt/odt system preserved negative ACES values through the rrt and preserved negative OCES values through the odt, but clipped them up to zero at the final stage on output to the display/projector
- Most recent experimental rrt/odt systems clip up to zero as a first step in the RRT (and also as a first step in the ODT)
- How should negative ACES values be handled?
- How are the rrt and odt tone curves extended below zero?
- We learned that negative pixel values times negative off-diagonal matrix terms yield highly visible artifacts. This required a special matrix multiply which zeroes all negative times negative occurrences.

Rec709 Negatives are Undefined

- Originally considered necessary (by some) for preserving horizontal edge ringing when digitizing analog video
- For Rec709, values below 64/1023 are undefined, especially for large image regions (often occuring from telecine)
- Common to find large regions at values of 20/1023 and 30/1023
- No common specified practices for how to handle this
- Even clipping up to 64/1023, is not ubiquitous
- If negative pixels are *lifted* above 64/1023 black, what is the transfer curve of the negative values? Is it reflected from that above 64/1023? Is it extended with the 4.5 slope near black for the 2.22 transfer characteristic? Is it 2.4 gamma reflected? Is it something else?

Questions about Negatives

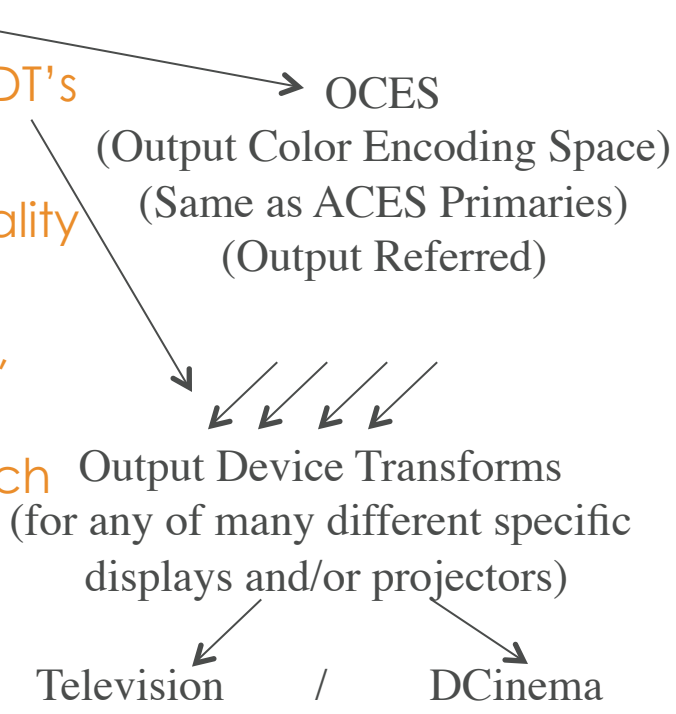
- Are Negative Values Necessary and/or Valuable in ACES?
- Should We Try To Preserve Negatives Through the ACES RRT/ODT System?
- Is It Worth The Substantial Extra Complexity in ACES?
- Can/Should We Do Anything About Values Below 64/1023 in Rec709?

No Reason The RRT Rendering Couldn't Also Be in Live Television Rendering

- ACES RRT/ODT rendering is providing a successful look in episodic television and motion picture production
- The more thorough rendering being standardized in the RRT/ODT system could provide a more pleasing look to some forms of live television (replacing current forms of Rec709 rendering)
- The RRT/ODT system is straightforward to implement as a 3D LUT (which is a commonly available system capability), and the full algorithms are within reach of current GPU processors

Common Codec?

- One common layered codec for dcinema and future television systems?
- Mastering to a single "look" for dcinema and television?
- How about mastering in ACES/OCES?
- Planning for multiple distribution paths via ODT's
- Fast track to HDR presentation
- Multiple layers, spatially, temporally, and quality (SNR)
- Common between production, contribution, and various distribution paths
- Eliminate encode-time binding of quality, such that quality matches display or projector capabilities via layer and ODT selection



Tuning 4k/8k Layered Coding

A

- Optimize for 2k
- Minimize bits at 2k
- More bits total for 4k and 8k

B

- Optimize for 4k
- More bits in the 2k layer
- Less bits total for 4k and 8k

- It takes more bits in a resolution-enhancing layer when a large number of bits are needed to correct lower resolution layers
- It is more efficient to code the lowest layers at high quality if planning for high layers
- Lower layers are still efficient, just not squeezed to "tolerable" artifact levels nor relying on artifact masking (e.g. not using deblocking filters)

Conclusion

- Television and motion picture technologies and appearance seem to be gradually merging
- Difference due to history are falling away due to improving practices and technology
- The hidden system architecture ingredient of the rendering transform is being made standard with the highest quality "Film Look" rendering
- Frame rates, bit depths, compression capabilities, color gamuts, display brightness ranges, and other key system features have recently been increasing and improving
- All the required system ingredients are working now, and can be demonstrated

Demonstration

SMPTE Meeting Presentation

The Unfolding Merger of Television and Movie Technology

Gary Demos

Image Essence LLC, Perris, CA, garyd@alumni.caltech.edu

Written for presentation at the

SMPTE Technical Conference

Abstract. *Live show video rendering to Rec709 typically has used a simple in-camera rendering. Movies until about the year 2000 were predominantly made on film, and printed to film.*

Movies have been predominantly 24fps, whereas live video has used higher motion rates.

The telecine, and also the digital film scanner, began to decouple the wide range film capture from the image produced for movie release or television release.

High-end digital cameras now capture a wide dynamic range, logically corresponding to the extended range captured on camera film negative.

A new system ingredient is the ACES system with a 16-bit half-float representation, a Reference Rendering Transform, followed by a device-specific Output Device Transform. This structure provides a unifying mechanism, bringing video and film technology closer together in underlying technology, and also in image appearance.

Keywords. Digital Cinema, HDTV, ACES, Mastering Formats, HDR

Introduction

The Dilemma of Standards Development

A problem in standards development related to the broad scope of issues being discussed here is the issue of parts of the standards already having been decided by one or more groups. This is sometimes referred to in international meetings as “decisions having been taken”. During the present author’s participation in the US HDTV standardization process during the period 1989-1996, paradoxically everything had already been agreed at all points in time, although everything was constantly changing during that entire time span. Thus, all contributions to the improvement of design attributes were treated to some degree as being unwelcome (although occasionally celebrated).

Although the current situation is much more open, there are still some issue of this type at present, such as the ITU BT.2020 specification of 7680x4320 and 3840x2160 as future HDTV formats. There are many other examples, but this serves as a representative example. The present author has ignored these issues then and now, since there is no alternative when attempting to bring all of the system attributes into harmony by considering them all in the environment wherein some attributes are already accepted (and decided) by some. Perhaps a justification for straying outside the lines of such agreement is that often not all stakeholders have been heard from, and the issues affect everyone who works with the technology. Another possible justification herein is the attempt to think of the system attributes and potential as a whole, rather than as incremental upgrades. It is believed that thinking broadly is more likely to stand the test of time, although it may be occasionally temporarily unpopular. At the current pace of technology advance, we are all humbled by an inability to think sufficiently broadly, anticipate new as-yet-unrevealed technical capabilities, and/or adjust as rapidly as may be needed. All we can do is to attempt our best effort in thinking things through.

It is also worth noting that former hardware and firmware is gradually and inexorably becoming software, including realtime software. Thus, the needs of signal specification are gradually becoming inherently much more malleable. That does not undo the need for standards and agreements, but what is needed is an ever changing landscape.

Further, we are at a stage where we are moving beyond good looking moving pictures to great looking moving pictures. Thus, it is becoming difficult to argue strongly against any approach or parameter set. It seems that we now aimed toward repurposing flexibility and preservation of as much image information as possible, which provides indirect value of potential significance. This is also an opportunity to address heretofore open issues, such as the lack of uniformity of appearance of television displays in the home (and elsewhere).

Origins of Digital Television and Digital Cinema

Video originated as an analog transmission system, including kinescope, and gradually expanded through the 20th century adding video recording and telecine. Beginning around 1985, video converted almost entirely to digital by 2000. Production and distribution of motion pictures for theatrical exhibition remained film-based throughout the entire 20th century. Production digital effects and digital animation for movies and for television gradually became significant over the last two and a half decades of the 20th century. For digital motion picture effects, digital film scanning and recording was used.

The first decade of the 21st century has seen a rapid advance in digital theatrical exhibition and digital camera capture for motion picture creation. Stereoscopic 3D theatrical exhibition has also re-emerged, now in digital form, with some deployment of 3D television as well.

Digital technology improves consistency, expands creative options, and enables increased quality when applied with care.

There have been fundamental differences in principles that have heretofore separated television and movie production, distribution, and exhibition. These principles have affected the delivered image in visually obvious ways, but these principles have their roots in the evolution of the fundamentally different underlying technologies.

In television, the reduced chroma resolution and chroma bandwidth of television YIQ, and later YUV, is often visible, especially at standard definition. The use of analog sharpening and noise reduction filters on the horizontal analog scanline is also often visible. Analog horizontal processing such as cable compensation and transmission filtering also can affect the image appearance. Interlace artifacts in capture, distribution, and presentation are often clearly (and objectionably) visible.

However, at present, there are numerous interlace-free television systems. These are sometimes called "progressive scan" due the historical analog scan, but are more correctly considered as being "whole frame" (especially when the camera opens a shutter to capture the whole frame at once, and the display or projector presents the whole frame at once, as with motion picture film). Although there was a strong historical legacy mindset which allowed interlace to survive the HDTV transition two decades ago, it is hoped that interlace is finally in its last years (although it will stay in interlaced images within the archives as a permanent issue).

Another fundamental issue separating television and motion pictures has been frame rate. The 24 frame per second (fps) update rate of motion pictures is insufficient for watching most sports. The 50 fps rate of PAL/Secam countries (and their HDTV versions) has been made interoperable by playing 24fps at 4% fast at 25fps (although this is a semitone higher in audio pitch and the increased pace is sometimes noticeable). The 60/59.94Hz rate of NTSC countries (and their HDTV versions) have used "3-2 pulldown" by holding even frames for two frames (or fields) and odd frames for three frames (or fields), or the odd/even variant. With interlace the issue of which field is "top" also comes into the configuration. Further, sometimes various video effects have been created at video rates, and composited with film elements such that various regions within each frame have different cadences and rates. A simple example is a credit scroll made on a video system at a video rate being superimposed upon (or composited over) a moving film image background. Fades, cross dissolves, and other simple common effects may also have this issue. Recreation of a 24fps release from such material is quite difficult, as is conversion to 25fps for 50Hz television systems. Thus 3-2 pulldown has long been highly problematic for repurposing, and remains so.

The Origin of Common Digital Hi Definition Parameters

The 1920 horizontal resolution of HDTV systems has its origin in a series of steps during the early conversion in the 1980's from analog to digital. In order to properly digitally sample analog television, a sampling rate of two or more times the 4.5MHz analog bandwidth is required. The color subcarrier of NTSC at 3.58MHz became another number of interest when sampling composite YIQ. A composite digital tape system using four times the 3.58MHz subcarrier frequency (called 4fsc) was used in some early digital video equipment, including digital tape recorders. Note that this is the origin of the "4" in the 4:2:2 nomenclature, although it is actually rare that there is an actual factor of 4 times anything, but rather the 4 is a loose (and probably poorly chosen) reference to complete sampling for luminance. The other two "2" values in the

4:2:2 nomenclature refer to the horizontal chroma color difference sampling of each of the two interlaced fields. Thus, 4:2:2 indicates half horizontal resolution for the U (R-Y) and V (B-Y) color difference channels. The only other two commonly used variants of this nomenclature are 4:2:0 and 4:4:4. The 4:2:0 format indicates half horizontal and vertical resolution of U and V and the 4:4:4 format indicates full horizontal and vertical resolution of U and V, or in the R and B channels of R, G, and B. For non-interlaced video, 4:4:4 or 4:2:0 make the most sense. For interlaced video, 4:2:0 has one U and V for two field times, hindering compression efficiency and causing additional interlace-related artifacts. Note that H264/AVC and MPEG-2 use 4:2:0 coding of interlace on Blu-Ray and in most HDTV distributions systems. There will occasionally be found a reference to 3:1:1 and other variants having reduced resolution, and also 22:11:11 or other ways of trying to say that HDTV is sampled at a high multiple of the NTSC 3.58MHz color subcarrier. However, these are becoming rare, probably because they are inaccurate descriptors based upon obsolete video concepts.

For component YUV (aka YPbPr and/or YCbCr) there was interest in having a single digital recorder which could record and playback 50Hz PAL and 59.94Hz NTSC in one unit. Given the everything-is-synchronized thinking at that time, a debate arose about choosing a factor times the least common multiple of the PAL and NTSC horizontal scanline rates around 15.7kHz. The least common multiple is 2.25MHz, and eventually the factor of six was chosen yielding 6 x 2.25MHz = 13.5MHz pixel sampling. Using YUV 4:2:2 sampling (discussed shortly) two samples provide color, yielding 2 x 13.5MHz = 27.0MHz. This rate, and various multiples of it, are prominent as a digital video system clock. For example, 2.75 x 27.0MHz = 74.25MHz, the most common HDTV pixel clock (and its 1000/1001 variant for 59.94Hz).

Having chosen a 13.5MHz pixel clock, a common number of pixels could be used for NTSC and PAL scanlines, including the unnecessary clocking of the retrace invisible pixels. This turned out to be 720 pixels (704 with a small edge pad). Doubling this number to make HDTV be double the resolution of this version of digital standard definition yields 1440 pixels. To expand the image aspect ratio using this width to go from 4:3 (1.33) of NTSC to 16:9 (1.78) of HDTV (more about this shortly), the 4/3 ratio is applied (since 16/9 = 4/3 squared). This yields the horizontal value of 1920.

At the time this was being debated, the present author pushed hard to alter this value to 2048, being a power of two and thus algorithmically more elegant for digital graphics computation. An independent push a decade ago for digital cinema chose 2048 for similar reasons, and thus today we have television systems at 1920 (and multiples thereof such as 3840 and 7680) and digital cinema presentation at 2048 and 4096. Much of this disparity relates to the image aspect ratio, and will be discussed shortly.

Returning to the origin of the 1920 x 1080 “common image format”, the 1080 height had its own set of steps. Initially the first analog HDTV system from Japan used 1035 interlaced lines. This was a little more than twice the 480 (or perhaps 483 or 486) lines of interlaced NTSC at 960, but was a little less than twice the 576 lines of interlaced 50Hz PAL at 1152. The 1920 number was chosen for this 1920x1035 digital interlaced system when moving from analog to digital.

The problems with interlace were raised concurrently with the problems of non-square pixel positions in the 720x486, 720x576, and 1920x1035 interlaced systems. The pixel spacings of these systems are 8/9, 16/15, and 23/24, respectively. As with the issues of interlace and the odd 1920 horizontal number, the problems associated with working with such slight non-square pixel spacings were also raised. However, unlike interlace and the value 1920, there was enough sympathy for the non-square pixel spacing that the vertical was changed to 1080, which yields square pixel spacing for 1920x1080 for the 16:9 aspect ratio.

The 720P Format

Since sports requires a higher frame than 24fps or 25fps, another debate arose around whether interlace was an enabler for high definition television, or just a bad idea. In the U.S., both Zenith and MIT proposed using a reduced resolution but without interlace. As mentioned previously, this was called “progressive scan” due to analog television scanning origins. Many interesting visual comparison demonstrations came from this work. Although several variations were also explored, such as 960-line interlace proposed by General Instrument and Sarnoff Labs, ultimately the discussion focused on a 720-line system (768 total lines including vertically invisible lines). Given the more computer-centric focus of the groups opposed to interlace, the square pixel spacing was also used, yielding 1280x720. Note that these numbers bear a nice 2/3 relationship to 1920x1080. They also bear a 4/9 relationship (2/3 squared) to the pixel count of 1920x1080.

In the mid 1990’s, the US decided to adopt both the 1920x1080 interlaced HDTV and 1280x720 non-interlaced system, which have thenceforth split the airwaves between them.

The Eighteen Formats

At the time this was being done, a list of eighteen variants were proposed for 1920x1080, 1280x720, and standard definition at 704x480. These included non-interlaced 24fps and 30fps 1920x1080, 1280x720, and 704x480, and also the 1000/1001 variants of all formats (based upon 59.94Hz vs 60.0Hz). Two opposing viewpoints formed, one claiming that eighteen formats is too many and is a burden, and a second viewpoint saying that any format that could be displayed should be allowed. This implied that as long as HDTV receiver manufacturers picked a set of formats they wanted to display, and signal provides (such as broadcasters) sent only those formats, that any set of formats could be used. There was further the confusion that the US HDTV standardization process was centered on terrestrial broadcast transmission and was essentially optional with respect to digital cable and satellite transmission, not to mention recorded media. The final standard struck out the eighteen format mandate. The practical affect of this was that both 24fps and 30fps 1920x1080 were not implemented by receiver manufacturers initially, and that 23.98, 29.97, and 59.94 rates were deployed to the exclusion of the 24.0, 30.0, and 60.0 rates. Thus, the de-facto result was three formats, being 1920x1080 59.94 interlaced, 1280x720 59.94 progressive, and 704x480 59.94 interlaced.

It was not until the Blu-Ray disk was deployed with HDMI supporting 24fps 1920x1080 that this very valuable one of the eighteen formats was actually deployed. This happened about a decade after initial HDTV deployment in the U.S. At present, many devices support the 24fps 1920x1080 format using the HDMI interface. However, due to this history, this 24fps format is still not present in systems such as U.S. Terrestrial HDTV, which allowed the initial group of receiver and decoder manufacturers and the initial broadcasters to bypass the 1920x1080 24fps format from the original eighteen that were proposed (and abandoned).

The HDTV and Movie Aspect Ratios

The 16:9 aspect ratio was recommended based upon it being in between the movie aspect ratios of 2.40:1, 1.85:1, and the standard-definition television aspect ratio of 1.33:1. A counter proposal was made based upon the premise that a wide screen format should be between the movie aspect ratios, and 2.0:1 was recommended. This had the further benefit of aligning the proposed 2048 horizontal number with a vertical 1024 resolution, making both be a power of two. The total pixels match the 1920x1080, but the aspect ratio is different. Note that 2.40:1

aspect ratio movies are usually high budget movies, which have historically had more resolution on film (nearly twice the film area vs. 1.85 images). With 16:9, a 2.40 movie has less resolution (via less height) than a 1.85 movie, contrary to the intent.

At present, wide screen HDTV is being discussed having an aspect ratio of 16/9 x 4/3 (yet again) yielding 64/27, which is 2.37:1, yielding 2560x1080. Note that 2560 horizontal, in the form of 2560x2048 (often cropped for 1.18 width to 2416 x 2048 for 2:1 squeezed 2.36:1), was a favorite film output format of the present author during 1981-1986. If this 64/27 aspect ratio comes about, then 16:9, at 1.78 (which is near 1.85:1), and this new wider aspect ratio near 2.40:1, will offer separate devices optimized for each (unless flexible aspect ratios come about, such as flexible display material on a roll, making it easily adjustable). Thus, the issue of 16:9 favoring 1.85 but not favoring 2.40 has come up yet again.

There is also a debate about resolutions above HDTV, being suggested at double and quadruple 1920, being 3840 and 7680. The same sort of legacy entrenchment against 4096 and 8192 seems to be present with respect to these numbers as was present with 1920 vs. 2048. It is not clear what the vertical aspect ratios of these ultra-high-definition systems will eventually be, but some of the early demonstrations have continued to use 16:9 (e.g. as defined in ITU BT.2020). Given that it has been nearly two decades since the 16:9 and 1920 choices were made, and given that we now have the benefit of hindsight, it might be an appropriate time to recognize that the reasons behind 2048 and the 2.0:1 aspect ratio recommendation were solid. Thus, perhaps there is now an opportunity to consider 4096x2048 and 8192x4096 as possible next steps instead of formats based upon 3840 nor 7680. This power-of-two-based path would also eliminate the need for the 64/27 aspect ratio idea with all of its implicit deployment complexity. Given that 4096x2160 is a DCinema format, this 1.9:1 aspect ratio outer container might also be a resonable choice, along with double resolution at 8192x4320.

Disappearing Anisotropy

With the waning of 4:2:2 (in favor of 4:4:4 and 4:2:0), the disappearance of interlace (hopefully!) and the concept of the scanline fading into history, the image raster is becoming gradually isotropic between horizontal and vertical. The move to 4:4:4 systems, and the use of 3 chip cameras and higher density color filter arrays (beyond the Bayer pattern), is gradually bringing red and blue resolution up to the resolution of green and/or luminance.

DCinema typically projects full 2k or 4k resolution for each of X, Y, and Z (as R, G, and B). New HDMI systems such as HDMI 1.4, and current dual-link and 3Gb HD-SDI can all support 4:4:4 resolution. Some home and professional HDTV displays are capable of 4:4:4 display.

The moving image raster is gradually becoming more uniform in resolution and color channels. Only a move from square to hexagonal pixel arrays remains as an unexplored optional way to become more directionally uniform, although computation is much more straightfoward using the current horizontal and vertical axes.

Compression Codecs

The U.S. HDTV terrestrial broadcast standard chose MPEG-2 as the compression codec. It was pointed out during the discussion that layered codecs are more robust and also inherently more flexible (although this was not widely acknowledged until later). It is easily seen in hindsight that not only does H.264/AVC represent a significant improvement over the deployed MPEG-2 HDTV Codec, but that future codecs will further improve beyond H.264/AVC, and can potentially also offer the benefits of layering.

With the Digital Cinema choice of the inherently layered JPEG-2000, a layered codec was selected which provides a base resolution layer of 2048x1080 and an enhancing layer of to 4096x2160. This represented a clear validation of layered compression. However, JPEG-2000 lacks motion compensation and is intended for high bitrates (low compression ratios). There are fundamental design weaknesses in all deployed distribution codecs such as the double-wide deadband quantizer, which hinder quality in JPEG-2000 (part 1) for DCinema, MPEG-2, and H.264/AVC. The double-wide quantization deadband is appropriate for eliminating bit waste due to high frequency noise (thus at full resolution), but is inappropriate when used at all detail levels as it is in current codecs. The layering in JPEG-2000 (part 1) for DCinema validates the layered robustness and feature discussion, but is implemented in a way that is statistically divergent (by using sub-bands without correcting outlier divergence, and due to the double-wide deadband used at all spatial frequencies).

The present author is demonstrating a layered error-bounding codec which yields high quality and high compression ratios sufficient to be a candidate unified cinema and television codec.

Frame Rates

As part of the U.S. HDTV system debate, alternate frame rates were proposed for consideration, such as 72.0Hz and 75.0Hz. These frame rates improved interoperability with 24.0Hz and 25.0Hz uses, and were flicker-free on the flicker-prone CRT displays of the early 1990's. However, as with the AC power cycle frequencies, each country tended to stay with their existing standard-definition rates.

Subsequent to the U.S. selection of its terrestrial TV system, other countries began selecting their own variants of the U.S. system, also built initially upon MPEG-2 digital moving image compression. The main difference was in European 50Hz PAL/Secam countries which adopted 50.0Hz interlaced 1920x1080 and 50.0Hz non-interlaced 1280x720 formats. The requirement for CRT power supply regulation to have a locked rate has been gone for over a decade. However, cameras seeing light flicker due to AC power frequencies remains an issue (although this can be mitigated sometimes by careful selection of camera shutter angle).

Also remaining as an issue is interoperation between 59.94Hz and 50.0Hz television signals. Standards converters between these rates cannot always function smoothly due to inherent issues in image temporal sampling. Even a 100% duty cycle frame is still a box filter which can temporally alias due to phase for repeating events at or near the frame rate.

One principle of conversion is that higher frame rates provide more temporal samples per unit time, and thus improve the conversion quality despite the imperfect temporal sampling. This does not remove captured (or synthesized) temporal aliasing, however, like backward spinning car wheels. These just have further temporal aliasing errors compounded when attempting frame rate conversion.

Since temporally sliced 3D left-right switching displays each eye's image at different times, odd artifacts occur when the eye follows a smoothly moving object. Such temporal 3D switching often occurs two or three times for each eye (e.g. 72Hz for each eye, being 144Hz for both eyes for 24fps movies). However, the redisplay of the same image, and then the switch to the next image, still yields artifacts when following motion. For these reasons, and due to the inherent inadequacy of 24fps for many fast motion scenes (like sports), there is now serious consideration of motion picture preparation and presentation at rates substantially higher than 24fps. The original Showscan rate of 60fps, and the Showscan tests of 72fps, are often cited, although these were film-based tests. With digital cameras and projectors, different temporal

properties (such as high shutter duty cycles in cameras), and improved steadiness and stable image brightness allow significant potential improvements.

The deployed JPEG-2000-based Digital Cinema Package (DCP) definition supports 48fps at 2048x1080, but not in 3D, and not at 4096x2160. There is also discussion of 60fps, which is not specified as an option in the DCP definition. Further, not all movie servers and projectors support this 48fps mode in the DCP definition.

Note that a high duty cycle for the open shutter, such as 90%, provides every other frame with a duty cycle at 45%, near the 50% to 60% (180degrees to 215degrees of rotation), which is the commonly used shutter duration proportion for 24fps. Many digital cameras are capable of high duty cycles.

The rate of 30fps has been explored for many decades due to its relationship to 60fps. The 30fps rate has found some use despite inconsistent support within equipment. A decade ago the present author explored 36fps at a 50% duty cycle using film. The 36fps rate has a natural relationship to 72fps. It was found that 36fps could utilize a cadence similar to 3-2 pulldown (2-1-2) with similar appearance on 60Hz. When presented on 72Hz, the 36fps rate retained some of the “story-telling” psychological distance of 24fps, but with noticeably improved motion rendition. Since digital moving image systems have inherent potential flexibility in image update rate, there is no reason that new image update rates couldn’t be supported directly in the way that 24fps is now supported over HDMI. The concept that the display is somehow always updating internally at 60Hz need no longer restrict image update rates, especially given that many displays actually operate at multiples of 60Hz such as 120Hz and 240Hz (or 50Hz equivalents). Even these higher rates are likely not necessary in that potentially arbitrary update rate could be supported up to some fairly high limit. With improved data rate compression systems and improved data delivery capacities on media and on wired and wireless networks, fairly high frame rates are also possible to convey. Given the discussion of 4k and 8k resolution, it is natural to also discuss a potential increase in temporal update rate at existing high definition resolutions as well as 4k and 8k.

It should be noted that there are a few moving image parameters that can be prominently perceived even with a very small image. These are color gamut, depth of black, brightness, and frame rate. Given that pixel motion in the frame rate range of 24fps to 60fps is often dozens of pixels per frame, an increase in resolution yields no clarity benefit to image definition for such moving objects. It is only when the motion stops or becomes very slow that increased resolution above high definition yields visible benefit.

For example at 60fps at a 50% shutter duration, an object moving the width of a 2k screen in 3seconds smears across 5 pixels horizontally (with 5 pixel gaps during shutter closed). This becomes 10 pixels of smear at 4k, and 20 pixels of smear at 8k.

Even at 120fps, these smears are still 2.5 pixels, 5 pixels, and 10 pixels across for 2k, 4k, and 8k respectively.

Flexible Display Synchronization

Video digital sync has been modeled on the scanning raster, with a locked pixel clock that is clocked out with specific counts for each scanline. The horizontal and vertical retrace pixels which are invisible, and arguably also irrelevant, are also clocked out. This is the format of HD-SDI, which carries the pixel clock as a small multiple of the 1.5gbps or 3.0gbps bit clock, and which clocks out the invisible retrace pixels. The HDMI cable similarly carries an embodiment of the 74.25MHz pixel clock (or its 1000/1001 variant).

While it is necessary for a camera to have a precise shutter-open time for a precise exposure, it is not necessary with most display and projector technologies to update the image at a precise rate. An image held for +-10% variation in presentation time for each frame will not vary in brightness at all. Only motion smoothness will affected. A variation of +-1% would likely not be perceptible even in motion smoothness.

Audio synchronization with the moving image requires that audio not advance more than approximately 15ms prior to image motion, and not lag more than approximately 30ms after image motion (note that SMPTE has ongoing work on these tolerances). Within these tolerances, however, is a great deal of inherent flexibility. Smooth audio should not vary in clock rate (usually 44.1kHz or 48kHz) and thus probably should use some form of locked sample clock. However, this is a much lower rate than the pixel clock. Note that SMPTE has been studying the use of GPS clocks for local and regional synchronization, which is expanding our existing knowledge about broadcast and facility synchronization systems.

If the image were to vary its rate over a tolerance relative to audio of 7ms before audio sync to 15ms after sync, the affect would likely be imperceptible (although the precise amounts of such potential variation should be thoroughly tested).

Note that common practice with today's synchronized displays, even on computers, is to drop or add a frame. This is a variation of 100% of a frame time! The frame time of 60fps is 16.7ms. An added or dropped frame at 60fps (thus 16.7ms) is clearly visible when presenting moving images. The frame time of 24fps is 41.7ms. This is both visible and objectionable, and is also outside of proper audio sync tolerance. Thus adding or dropping a frame at 24fps should be avoided (although it is common, especially on computer screens, hand-held screens, and note-pad screens).

Perhaps a good choice is to allow no more variation than +-3% image update synchronization flexibility, while seeking to maintain audio to image synchronization within the range of 3.5ms advance to 7ms retard of the image with respect to the audio.

Given this tolerance range, and given a typical modern display or projector which does not vary in brightness with image hold time nor image update pace, it is possible to architect image presentation which is indistinguishable during viewing from the inflexible locked-clock video systems. In the modern flexible data world of digital networks and wireless data transmission, such flexibility is likely to yield a significant presentation quality improvement over the presently deployed frame drop/repeat. The lack of a locked pixel clock also is inherently easier when utilizing commonly available (but data-centric) connection technologies (e.g. infiniband).

Note that unlocking the pixel clock does require that some software (or firmware) algorithmic phase-lock-loop equivalent exist, albeit with a very loose lock by digital standards (being in the milliseconds). Note that a millisecond in a GHz-class digital system is a million clocks! The exception is the need for speed-of-light considerations (and possibly also packet-retry considerations) in networks, which can also yield millisecond (or longer) variation. Usually this is dealt with by buffering, although buffering yields corresponding latency when there is interaction.

It is therefore feasible, although as yet undemonstrated, to provide the motion and audio synchronization quality of a pixel-locked system (e.g. and Blu-ray player), on broadly deployed decoupled moving image players where pixel clock locking is difficult or infeasible.

Bit Depth

Deployed distribution compression formats have heretofore been limited to 8bits (e.g. deployed MPEG-2 and H.264/AVC). Even less than 8bits are actually delivered due to quantization, especially at lower bitrates. Further, some codec features, such as deblocking filters, make no attempt at coding accuracy, being intended solely to mask or suppress visible codec artifacts. With commonly used 16-bit signed integer computation there is an effective limit of about 10 ½ bits for codec precision. The conveying of high quality compressed images at higher bit depths is possible with higher precision computation (such as 32-bit floating point computation). The dynamic range of camera capture is usually beyond the range of presentation. Further, the dynamic range of presentation can vary significantly. The use of 8bits in distribution codecs implies limited dynamic range in the mastered moving image as well as limited dynamic range on the distribution presentation display. These are principles inherent in the U.S. HDTV system, and are implicit in Rec709 as a signal format for HDTV. Alternatively, the Digital Cinema format specifies 12-bits (although computation using 16-bit integer implementations are prevalent, and XYZ is used which is perhaps ¾ of a bit less efficient, giving an actual RGB equivalent limit for DCinema, even at maximum compressed bitrates, of about 10 bits). Further, Digital Cinema system parameters are intended for 14fl (which is 48nits noting that cd/m^2=nits, using nits henceforth) dark-surround projected presentation, such that higher dynamic range and brightness will need additional bit depth (beyond the 12 bits needed for imperceptible lsb transitions at 14fl, which is 48nits).

Note that quasi-log and gamma representations are optimized for specific dynamic ranges and sometimes for specific absolute brightness levels. This contrasts with pure-log, and floating-point, which are in some sense range-independent. Floating point is logarithmic in the exponent field, and linear in the mantissa field (over each range of two, being one stop for pixel light). Floating point has the advantage of supporting negative numbers, which are not available with a pure logarithmic representation. Negative numbers will be discussed below.

Screen Brightness

Movies have long been presented in a dark surrounding at a peak white level of 10fl to 14fl (34 to 48nits). The television mastering brightness has historically been about 20fl to 25fl (68 to 85nits) on reference CRT displays. Flat panel displays, both LCD and Plasma, have often been used in recent years at much higher brightness in both mastering and final presentation. Further, some displays have deep blacks, such as the OLED and also regionally-modulated LED backlight LCD displays. Tonal distinction also naturally increases with these technologies to higher bit depths over any given dynamic range.

Although it is not simple to say how much increase in dynamic range corresponds to how much increase in needed bit depth, we do know some things. It is simple to test a gamma, such as gamma 2.4 for presentation of Rec709, or gamma 2.6 for DCinema, for the step size of a least significant bit (lsb). However, it is easiest to think in terms of a pure logarithmic representation, wherein each step up is the same percentage increase up, independent of the value at that step. Using quarter-percent steps as an approximate perceptual discrimination limit, the number of bits needed to represent any given dynamic range is fully determined. However, this simple approach is usually complicated somewhat by optimizing the dark and bright regions for a specific dynamic range, and perhaps a specific absolute brightness for the representation. The most common “log-like” representation is a simple quasi-log, usually of the form log(x+a) where the “a” term is a constant. Another common simple quasi-log modification is a slope-matched straight line from a low value (e.g. 1% of white) down to pure black. Usually these modified log representations optimize less than a bit around mid-tones (10% to 20%), but can significantly modify the size of quantization steps at low and high brightness.

The use of the 16-bit half float, as used for exr files and for ACES, adds a bit for representing signed numbers, and allocates the remaining bits as 5 exponent bits and 10 mantissa bits. When representing linear (unity gamma) light, as with ACES, this can be considered to be 32-stops of range (for the 5 bits), and 1024 code values per stop. Note that the 32-stops are a logarithm, and the 1024 code values are a piecewise-linear interpolation in-between the 32-stops of logarithm. Note further that 1024 code values per stop, a stop being a factor of two range, is between about 0.025% at 0.05% in self-relative steps (the self-relative step size is a factor of two higher for low values of the mantissa than high values at a given exponent, since the magnitude meaning of the values themselves are a factor of two different whereas the steps are the same absolute difference). Another way to view this precision is that ten concatenated computations with error at +-1/2 lsb for each step will have maximum outliers at 10 x .5 x .0005 = .0025, which are quarter-percent steps. Note that the system rendering gamma boost (discussed below in the RRT section) can double this step size, such that a maximum of five concatenated steps is a safer choice.

Ambient Surround and White Point

Digital Cinema has a dark surround, whereas typical television viewing may have dark surround for evening viewing with low room illumination, but high ambient surround if room lights are bright, and/or if daylight is coming into the room. The scene darks and lights are perceived differently, as are the saturation levels of colors, depending upon the ambient surround. The white balance may also be perceived differently if the color of the ambient surround changes or is different from that intended. Note that Rec709 uses a D65 whitepoint when R=G=B, whereas DCinema uses a greenish whitepoint near D60 when R=G=B for calibration, but allows any whitepoint in the scene in the X'Y'Z' representation used for the DCP. For example warm scenes might be balanced for D40, and cool scenes for D70, with normal scenes around D55 or D60. Note that the "whitepoint gamut" allows full brightness over the range D55 to D65. With a dark surround in DCinema, it is possible to present a range of whitepoints. With ambient surround light present near the display, however, the white perceived from the display is affected significantly by the nearby surround light color.

Color Primaries

The initial NTSC RGB primaries had a well-chosen gamut. However, the color primaries for NTSC changed with time, primarily to match the easily-available phosphors by switching to SMPTE-C primaries with a reduced gamut. Rec709 expanded the gamut for HDTV. DCinema selected the "minimum gamut" defined by the "P3" primaries. DCinema chose CIE 1931 XYZ (at gamma 2.6) for the color space.

All of these gamuts are defined using CIE 1931 x_bar, y_bar, z_bar color matching functions for 2degree matching, based upon an average viewer. Actual viewers usually differ somewhat from this average, and image regions for a given color often are larger or smaller than 2degrees in the viewing angle they subtend. The CIE 1971 10-degree color matching functions describe larger color region matching, and the CIE 2006 parametric color matching functions represent a model for the range of 1degree to 10degrees, and further model average age differences in color perception over the range of 20 years to 80 years old. Note that narrower primaries for wider gamuts cause an increase in individual person variation vs. all of these color matching systems based upon multi-person averages.

Some affordable displays can achieve the P3 gamut using OLED and LED-backlight LCD technology. Laser light sources can extend the gamut for projection, allowing one or more

primaries to be located on the spectrum locus if desired. Displays and projectors can add additional primaries beyond R, G, and B, which can help both expand the gamut as well as reduce individual color perception variation (by broadening the displayed spectrum).

The computational burden of modeling the spectral color matching functions, and even individual variation and region size variation using the parametric CIE 2006 color matching functions, is now well within the power of the GPU in most computers. Further, the "color appearance models" which describe and compensate for white point alteration and ambient surround are also becoming feasible.

Output Referred Rec 709

The Rec709 definition includes a gamma 2.22 function (as the inverse of .045) with a slope-matched straight line near black. This was described as being a camera "transfer characteristic" (also called an "electro optical transfer function"). However, it was eventually recognized that the display gamma was not defined by this. In the last couple of years a 2.4 gamma for display was defined. By defining the display gamma of Rec709, the specification becomes "output referred". The Rec709 specification is actually not simultaneously "scene referred" from the 2.2 gamma, since common practice involves a gamma boost and other adjustments with respect to scene light (discussed shortly). Thus, the 2.22 gamma (with a straight line near black) is a relatively useless specification in practice.

The Rec709 specification defines CIE 1931 xy chromaticity coordinates for the R, G, and B primaries and specifies a D65 whitepoint as chromaticity. The Rec709 specification also defines the YUV (aka YPrPb digital or YCrCb analog) transformation to and from RGB. This YUV transformation is performed as a linear matrix operation on the non-linear gamma-adjusted signals, and thus is actually a non-linear operation (which has a complex meaning since a matrix is a linear operator). The operation is invertible for 4:4:4 (possibly within one or two lsb precision, depending on implementation), but is not invertible for 4:2:2, 4:2:0, or other filtered image formats due to the reduced resolution inherent in these subsampled formats.

Conceptually a Rec709 signal is produced by a camera, and can be directly plugged into a display to provide an output-referred ready-to-view image. Thus, there is an implicit conversion from the light in the scene to a pleasing viewable image which is performed within the video camera. This is often done with adjustments possible within the camera, often adjusted remotely. Such adjustments allow "painting" of the scene to occur between the light enterring the lens and the Rec709 output. If such painting is done in a switcher or other downstream processor, then a neutral Rec709 output is typically provided by the camera, but such neutral Rec709 output is still rendered as an output-referred image, albeit an intermediate rendering. The headroom for bright parts of the scene and the ability to lift the darks is usually quite limited in this partially rendered Rec709 form. Often switchers are limited in their function to wipes, dissolves, fades, and other simple processing due to this limitation.

The conversion from scene-referred light to output-referred light is called by many names in color science literature, but is perhaps most commonly called "rendering". This differs from the "rendering" wherein a computer generated image is converted from shape descriptors, surface description, and a lighting environment description into viewable pixels. The rendering process happens within a camera when the camera is outputting Rec709 video. This "in-camera" rendering is required for video cameras which produce a Rec709 output, which is essentially every high definition camera.

The major processing steps that perform this rendering are typically a matrix process, a gamma boost, a dark adjustment, and a highlight adjustment (usually called a "knee"). The matrix and

gamma together affect color saturation, and may do so in a way that affects different colors at different brightnesses differently.

Still Cameras and Color Paper Prints

The still image camera can also utilize an intermediate format which is not rendered or is partially rendered, but more typically outputs a rendered image ready for display. A separate process converts the intermediate format or rendered image for color paper printing by providing proper conversions (and usually some additional rendering) for dyes printed on paper.

Use of a camera-vendor-specific "raw" format, when available, usually allows direct access to the scene-referred (and thus un-rendered) image. Rendering is then done in tools like Photoshop or Aperture. The ACES system opens this methodology up using the device-independent standardized scene-referred half-float ACES file format.

The Film Model

The film model is inherently different. In the film model, a color negative captured scene light into a density representation of yellow, cyan, and magenta dyes. These dyes are then printed, usually with some exposure and color balance correction (called "printer points") onto a positive color film print. The negative contains an extended dynamic range of exposure beyond that of the print, such that usually one to three stops of up and down exposure adjustment is available. The negative-to-print process renders scene light into projected light at the theater. Additional rendering adjustments are possible by "flashing" to reduce contrast and boost and lift dark detail, "bleach bypass" to alter color, and "push" or "pull" exposure processing, etc., or simply by choosing a specific type of camera negative and/or print film.

Prior to HDTV, more sophisticated high resolution film-based rendering involved either digital scanning and recording, or optical red, green, blue separation film processing. Sophisticated processing such as blue-screen or green-screen at high resolution can be processed digitally or via multiple black-and-white color separation optical film steps.

Once HDTV came into existence, the HD telecine enabled color rendering, blue-screen, and other sophisticated processing. This was initially done in multiple units, at first using analog processing, but later with realtime digital processing. Eventually the HD color corrector embodied the functions of many formerly separate units into the single color corrector unit.

Note that color rendering was embedded as one of these processes. Also note that this architecture allowed different photography, under different lighting conditions, to be brought into a common appearance for compositing. This often involves different exposure and color balance, but sometimes different gammas, different color saturation levels, or other elaborate rendering appearance alterations to bring the composited elements together and provide them with the desired overall appearance.

Benefits of Higher Dynamic Range and Scene Referral

The extended dynamic range of film negative has long been proven beneficial in production, post-production, telecine mastering and re-mastering, and archives. At present the extended range of digital cameras offers these same benefits. The quasi-log camera data pixel representations, beginning with Thomson/GV "Filmstream" for the Viper, have proven directly useful with color correctors. The Filmstream quasi-log (in the form of log(x+const)) format was a

simple transformation from the linear-light seen by the sensor, with almost no other processing. However, by bypassing these quasi-log formats, or by directly utilizing linear (unity gamma) sensor data, the sensor data can be directly utilized. Given a transform to the linear (unity gamma) ACES primaries using an Input Device Transform (IDT) process, an OpenExr 16-bit half-float representation can provide this same close-to-what-the-sensor-saw representation. This scene referred information becomes common across different cameras when brought into a common color representation such as ACES (by using the ACES RGB primaries). If an archival digital medium is developed, the archive of ACES data should contain everything the sensor saw, and would thus be essentially equivalent to storing film negative. All of the film negative advantages are embodied in the ACES model.

Reference Rendering Transform

Recall that rendering (in the meaning being used in this paper) is the bridge between scene-referred pixel data (as seen by the camera sensor) and output-referred pixel data (as displayed or projected).

Heretofore the rendering of scene-referred images has usually been performed to a specific device, such as a color paper print. Rendering to an intermediate format is also sometimes utilized, which provides the benefit of some level of device independence.

Rendering has often been hidden, and is sometimes proprietary and undisclosed. Rendering sometimes comes within devices such as cameras, color printers, or software tools.

As mentioned above for cameras, rendering typically consists of a gamma boost, a color matrix (which may embody a saturation adjustment to some colors), a highlight knee process, and often some adjustment near black. This structure is predominant, no matter how the rendering is embodied, as a single whole process, or as pieces of a process using an intermediate file or datastream. In the film embodiment, the matrix could be considered as being replaced by inter-layer affects, both optical and chemical. Since the initial digital film scanning systems in the 1970's, including the present author's, a 3D LUT having cross-primary-color terms has often been the practical implementation embodiment of the film rendering. This is usually implemented as a process between the scanned pixel densities (as printing density, or as some derivation of status-M density) to the pixel values being sent to a film recorder. A model of the linear light in computer generated images (CGI) and how it is transformed to the film recorder pixels was also needed. However, with the advent of digital cinema projection, and with the need since the original systems in the 1970's to have a color monitor during working, a full model of the displayed and/or projected light is and was always required. Thus, a full model of negative film's embodiment of scene light, and the negative-through-print implicit rendering, as well as the light exposing the film print through to light (perhaps via printing density) was all required from the initial development of these systems in the 1970's.

The ACES system is the first attempt to standardize the full breadth of rendering being discussed herein. Given the ambitious nature of ACES, and the breadth of technology that it embodies, there will continue to be some refinement. However, a working system for ACES is now in place, and it is being used successfully in production. Various key components of the ACES system are being standardized within SMPTE.

The reference rendering transform (RRT) is being developed to be the standardized ACES conversion from the ACES scene-referred pixel values (usually as half-floats) in ACES RGB primaries, into output-referred pixels in the Output Color Encoding Space (OCES), which also uses the ACES RGB primaries. The OCES pixels can then be given an Output Device

Transform (ODT) for a given device having defined colorimetry, transfer function (e.g. gamma), and dynamic range.

The ACES Neutral Spine

The ACES whitepoint is D60. This is the xy chromaticity when R=G=B in ACES primaries. The ACES primaries are partly non-physical, being outside the spectrum locus for the green and blue primaries. The blue primary is also slightly negative in y chromaticity.

ACES is architected such that the RRT transforms R=G=B in scene-referred ACES into R=G=B for output-referred OCES. This transformation can be considered a grey or "neutral" mapping, and is sometimes called the RRT "tone curve". When the reference ODT (also called the Reference Device Transform, or RDT) is included, the combined RRT + ODT tone curve is sometimes called the "system tone curve". The "system tone curve" is specific to one ODT, and usually refers to the DCinema ODT for 14fl (48nits) projected images having gamma 2.6 with P3 primaries with a D60 whitepoint. Note that X'Y'Z' does not have X=Y=Z for D60 white. The maximum value of D60 tristimulus in the 0.0 to 1.0 range is X=.944312344, Y=.991252018, Z=1.0, and for D65 (e.g. as in Rec709) is X=.873843017, Y=.919553073, Z=1.0.

The neutral spine is altered primarily by a gamma boost in the neutral RRT tone curve, but the slope of this curve is a function that extends from black to well over 100.0 times reference white (actually as far as the half-float or float numbers will go). The gamma boost is highest near 18% input of reference white at 0.18, which maps to 10% output of reference white at 0.10. This is based directly on the Laboratory Aim Density (LAD) system for film negative and film print. The slope of the RRT tone curve varies starting with a shallow slope near black, up to a peak slope near LAD, then gradually becoming more shallow up to very high values.

Note that rendering transforms typically have not mapped neutral to neutral, but have tended to map neutral to various off-white colors at various points in the tone curve. For example, a print film emulation will typically map neutral scene highlights to a yellowish tint corresponding to the yellowish tint of clear print film.

The mapping of D60 to D60 in the RRT was an excellent design choice in the ACES architecture, which greatly simplifies the system, as well as providing a key principle for arbitrarily extending dynamic range (both dark and light).

In order to render while retaining a D60 to D60 mapping, there must be fairly strong adjustments moving gradually away from neutral. Face tones, for example, must adjust strongly even when they are moderately close to neutral. In working on the RRT, this has been a sensitive region requiring care to ensure local smoothness and sufficient precision while still attaining the desired face tone color rendering.

The RRT Development History

Initial RRT designs centered on a density model for film negative and film print. The film density RRT was utilized and explored between 2006 and 2011. A key development during this period was the modeling of colors and brightnesses beyond the range possible with film-based dye densities. Film prints model red, green, and blue as pair combinations of dye layers as yellow + magenta, yellow + cyan, and cyan + magenta, respectively. Since these are two layer combinations, the maximum transmission of light in red green and blue is much less than the amounts of red, green, and blue which can be made with additive display and projection technologies.

The ability to extend dynamic range, beyond the logical limit of clear film (minimum density) with 14fl (48nits) projection, was also enabled and added by extending the design.

During 2011, a simpler empirical model more like that used in video cameras was begun. Many of the basic film-density (which is negative logarithmic) concepts were adapted to the additive linear empirical design. This approach has been continuously refined and improved up to the present, and is currently in use. The current RRT/ODT system is technically and mathematically sound. Additional technical ingredients, some of which are discussed here, are showing significant value, and further refinement and augmentation of the RRT model is being actively pursued.

The goal is to continue refinement of the RRT/ODT technical ingredients to the highest level of quality and utility, and once finalized submit them for standardization.

ODT Development History

Output Device Transform (ODT) development corresponded approximately to RRT development, with new ODT's being explored concurrent with new RRT's. The emphasis is on DCinema, but gradually the broader landscape of useful output devices began to also be explored and supported. The emphasis on DCinema makes the ODT for the P3 color gamut central. This became known as the "Reference Device Transform" (RDT).

Also high dynamic range (HDR) displays and projectors require a corresponding HDR ODT.

Device Primaries

Note that there is an inherent issue in how to reduce color gamut.

For Rec709, the reduced gamut of primaries, and how to map out-of-gamut colors into this range, has remained an issue to be further explored (although there has been some exploration within the Rec709 ODT). With Rec709 television systems, the limited gamut has inherently avoided most gamut adjustment issues. However, many home television displays have significantly wider gamut capabilities than Rec709, raising issues about how Rec709 should/could be extended, and leading to YCC as a possible gamut extension mechanism (although testing and deployment has been heretofore limited). ACES encompasses the full gamut by design, avoiding the need to explore extension.

For color gamuts wider than P3 with narrow primaries (e.g. some laser projectors), the issue of extending, testing, and exploring the gamut for ODT's remains, as well as compressing a gamut wider than P3 into P3's gamut and Rec709's gamut. Further, it is unclear (perhaps even unlikely) that CIE 1931 will be adequate in this case, thus potentially requiring some future extensions or modifications to the ACES architecture beyond just the ODT (since ACES is CIE 1931-based, as is Rec709).

Functions Within The ODT

The High Dynamic Range (HDR) Output Device Transforms (ODT's) intend to present a larger output range, corresponding to the very wide range of ACES and the very wide range RRT output OCES.

The ODT must serve a number of image processing functions. These include remapping wider dynamic range in ACES to lower ranges by compacting details for shadows and highlights, mapping into the required gamut (if reduced), and white-point adjustment, as needed.

It has been recently suggested that the ODT system should perhaps be considered in terms of processing components. A first component might correspond to specific approximate dynamic ranges, or might be an adjustable algorithm which could be set to the appropriate intended output dynamic range (although some of this function might also be found within the final stages of the RRT). An essential component for every ODT is to apply device-specific transforms yielding signal types needed for that device type, having that inherent approximate dynamic range. This essential ODT component thus would apply no appearance adjustment, but rather would represent only a translation of the desired appearance output of the first ODT process into the correct signal format needed to put that image appearance on a given display.

For limited range devices, such as Rec709 displays and 14fl DCinema (48nits), a tone curve with an S-Shape would be used in the first component to map the OCES extended range into the needed reduced range. For higher dynamic range presentation, such as using the Dolby Professional Reference Monitor in “Dynamic” brightness mode having a white at 600nits (175fl), a corresponding HDR first component would be utilized, perhaps with an asymptote for highlights instead of an S-Curve. This will be further discussed shortly.

For gamut reduction such as OCES images covering a P3 gamut being presented on a Rec709 display, it is unclear whether the first ODT process or second ODT process, or both, should be involved. For devices which cover the intended gamut, this split-ODT approach would suggest that the second ODT process would contain no appearance adjustment.

As yet unexplored is how an image appearance model, such as applying adaptation to adjust the image for a given ambient surround, should be handled. It is likely that both the first and second ODT process portions, and perhaps another in between them, might need new appearance modeling ingredients. The one example that we have experience with is the whitepoint alteration of ACES D60 to Rec709 D65, wherein we chose to support both D60 to D65 (via Bradford or CAT02) and D60 as D65 (without corresponding whitepoint alteration). In this example case, both approaches have proven heretofore useful.

Note that the two primary dominating terms in color appearance modeling are the ambient surround and the whitepoint adjustment. There are other significant terms, however, such as absolute presentation brightness, what scene precedes another (e.g. after-image affects), etc. The field of color appearance modeling is still a broad research topic, and much of image presentation’s potential range such as wide gamut and high dynamic range has only been sparsely explored. Note also that there are some significant image affects, such as color alteration and black lift when viewing a display from off to the side, which cannot be reasonably modeled nor corrected (at least when there are multiple viewers at that display, in this example). Although there is much that can be improved with color appearance, some such issues will likely intentionally remain unaddressed.

RRT Static Matrix

A typical static-matrix, like that used within the common in-camera rendering model, has been explored and used successfully. The matrix rows (or columns, the three terms for each matrix output of r,g, and b) all sum to 1.0, preserving r=g=b=D60 through the matrix. The matrix is applied to the linear (unity gamma) ACES values.

Saturation Control

A dynamic matrix is being explored which reduces off-diagonal matrix terms going away from neutral. The off-diagonal positive terms limit color gamut, but are needed near neutral to correct face tones and other critical colors. The off-diagonal negative terms in the blue-matrix row boost saturation of red, orange, yellow, and green near neutral, but are not needed for saturated ACES colors. Thus, the matrix is dynamically adjusted as a function of saturation, with renormalization after the off-diagonal terms are reduced.

Another additional feature being explored empirically, modeled upon film appearance, is the brightening and color saturation boost of face shadows. This was a feature in the film model which is re-created using hue-region processing in the red and yellow regions.

A function using 90-degree (slow-in) or 180-dgree (slow-in-out) cosines is used to ramp-in functions going away from neutral without slope discontinuities (which can create visible mach bands). This is especially needed crossing neutral from one hue region to another due to distinct independent processing within each hue region.

The RRT preserves the purity of saturated colors due to these ingredients, while maintaining the subtle but rapidly changing tones in the region between neutral and face tones. Face shadows are warmed and given a small saturation boost, similar to film.

RRT Tone Curve

An RRT tone curve is utilized to model film's appearance and implicit gamma boost and LAD-mapping behavior within the dynamic range of film, but extended to arbitrary dynamic range.

A question arises concerning how to apply this tone curve to ACES values. Should it be applied before or after the matrix? Both have been explored, and both have shown that they can yield good results, although they clearly behave differently.

More significantly has been whether to allow red, green, and blue values to be applied independently through the RRT tone curve, or whether to restore the r, g, and b ratios and do the RRT tone curve with a norm. With independent lookup, clearly the ratios are not preserved, and the ratios are therefore altered as a function of where each of r, g, and b fall on the tone curve. This can have the affect of moving reds toward orange and yellow in the shallow slope of the bright regions of the RRT tone curve, for example.

Several types of norms have been explored. One of the most successful has been $(r^2 + g^2 + b^2)/(r+g+b)$. Note that luminance was explored and rejected, in that blue looks up on a region of the tone curve a factor of six away from green (and red a factor of two from green), resulting in undesirable tone curve mapping.

The ratio-restore is performed relative to the norm, and is given a gamma boost which boost saturation. The amount of gamma boost can further be adjusted so that there is a higher gamma boost for desaturated colors (e.g. 1.5875 near D60) and a slightly lower gamma boost for saturated colors (e.g. 1.45). Note that this mechanism applies gamma boost as a function of saturation, but this gamma adjustment is otherwise invariant for all ACES levels (although ACES brightness is intentionally altered, using the norm, by the RRT tone curve). Thus, this gamma-boost aspect of RRT rendering yields the same color adjustment for every color, independent of scene exposure setting. Another way of thinking about this is that chromaticities are moved in a defined way away from D60 to boost saturation.

The use of a ratio-restore with saturation-adjusted gamma boost relative to norm lookup with the RRT tone curve affords the greatest control.

Note that negative values can be an issue if negative ACES values are processed in the RRT, although current RRT's clip ACES values up to zero as a first step. The negative terms in the blue-row of the matrix can be an issue for positive reds, oranges, yellows, and greens if the tone curve is applied after the matrix. The resulting negative blue value must be interpreted in the norm, as well as being interpreted on the tone curve. Should this, for example, result in a negative value of the norm, or should it adjust the norm value upward or downward? If the norm is negative, what should the corresponding negative region of the tone curve look like? If the blue is negative via ratio-restore or from the matrix, how should it be handled in subsequent hue-region and other processing in the RRT and ODT? Negative ACES values, and negative color primary values within the RRT/ODT system are discussed in more detail below.

Hue Region Curves

A key function of color rendering between scene referred and output referred images is to correct for over-bright presentation of some of the saturated colors which inherently occurs at some brightness levels. Initial empirical red, yellow, and blue hue region designs adjusted contrast as a function of color saturation. More recent designs have moved to utilizing a curve to adjust brightness as a function of color saturation and brightness within these hue regions. The red, yellow, and blue regions are blended using a cosine-based smooth overlap as a function of hue angle. This allows a film-modeling behavior to achieve desirable film-look appearance, albeit extended for high dynamic range.

The Normal Range ODT First Stage

For normal (e.g. Rec709 or 14fl=48nits) dynamic range, an ODT tone curve is applied to reduce the extended OCES range into the available darks and whites within that range. In a linear vs. linear plot, this ODT curve looks like an S-Curve, as might be expected. This serves a function somewhat similar to the highlight "knee" as well as the dark adjustments used by in-camera rendering for Rec709 output in video cameras.

As with the RRT, the issue of whether to look up each of OCES r, g, and b independently, or use a common norm with a ratio restore, have both been explored successfully. It is felt that greater control is available with the ratio-restore and norm. It is to some degree by accident if hue and saturation modifications inherent in independent r, g, and b lookup yield the intended appearance (although it does sometimes happen). For example, the highlight portion of the normal range ODT tone curve can also be useful as a natural and smoothly varying desaturated highlight. Alternatively, the norm which is looked up on the tone curve represents a D60 black- and white fully desaturated image element which has also proven useful when applied to highlights.

Preservation of bright saturated colors with smooth behavior to darker and less saturated regions, however, requires a mechanism involving ratio-restoration or preservation. This must be computed to provide some naturally-appearing color in bright saturated regions without exceeding the maximum output device range for any of r, g, or b. A common example in testing has been the preservation of bright car tail-lights or signal-lights while not injecting too much color into face highlights, and other desaturated highlights within the scene.

Processing often includes weighted blends of various image processing stages, and usually includes use of the slow_in_out and two-term-norm functions to smooth the color space in the highlights.

Note that with film all over-exposure goes to the minimum density (dmin) of the print film, which is a slightly yellowish off-white. The ODTs desaturate to d60 neutral, but otherwise retain some of the desaturation behavior of film highlights, but blended with some saturation preservation for bright saturated colors.

The First Stage of the HDR ODTs

The HDR ODT first stage usually uses a very different approach than normal dynamic range ODT's, although all HDR ODTs are still somewhat experimental. The typical HDR ODT attempts to minimize the ODT processing, eliminating the ODT tone curve and replacing the highlight processing with an extensible asymptote. This simple model allows a more direct view of the RRT, by minimizing the ODT, while retaining much of the appearance of normal range ODT processing.

The asymptote and other ingredients of this simplified HDR ODT design are inherently adjustable and extensible over a variety of high dynamic ranges, with only a few parameter alterations. Thus the first stage of the HDR ODT might best be constructed as a single ODT design, but one which is parametric as a function of output device dynamic range.

Applying ASC_CDL to ACES

The American Society of Cinematographers defined the Color Decision List (ASC_CDL) as a simple set of operations which allow useful adjustments. The space in which these are applied has remained unspecified, and reproducible results require also knowing what space was utilized in any given instance.

The ASC_CDL 1.2 operations are first r, g, and b scale factors, then r, g, and b offsets, followed then by an r, g, and b gamma, followed by a saturation adjustment.

There is a general sentiment that ASC_CDL should be applied to ACES in a single standardized way. There are advantages to applying ASC_CDL in the Look Modification Transform (LMT) position, which is applied to ACES scene-referred linear (gamma 1.0) light prior to being sent to the RRT. One could also make a case for applying ASC_CDL at the OCES position, between the RRT and ODT. However, we presently only have a small number of examples of high dynamic range displays, and so we are still early in our understanding of the higher brightness portions of OCES. The recommendation here is to apply the ASC_CDL in the LMT position to ACES prior to being sent to the RRT.

The r, g, and b scale factors of ASC_CDL then become equivalent to adjusting the lens and/or shutter. A factor of two is one stop brighter, and a factor of half is one stop dimmer, when applied in ACES, which is scene-referred light. The r, g, b offsets are similar to overall flare correction, albeit with no regional capability. The gamma adjustments for r, g, and b are probably typically used as a single common adjustment, to adjust appearance gamma in series with the implicit gamma boost within the RRT. The saturation operator is still being discussed. In its current ASC_CDL 1.2 version, the color moves toward r=g=b using a luminance equation, at saturation value 0.0. At saturation value 1.0, there is no adjustment, and at values above 1.0, saturation moves away from luminance. The saturation boost done from lumimance in this way boosts brightness of red and blue more than green, which can appear unnatural for large saturation boosts. There is also active discussion about possibly having additional adjustment parameters for saturation. Thus further exploration and refinement of the saturation operation is needed, and is actively being pursued.

Applying ASC_CDL using ACES r, g, and b primaries has felt workable but somewhat unfamiliar, in that ACES green is somewhat to the cyan from the commonly used green primaries. The red and blue primaries are also a little different than the behavior of red and blue as commonly expected. It is therefore also recommended that ASC_CDL be applied to an ACES value which is transformed using a matrix to P3_D60 primaries. The P3 r, g, and b primaries all move in a comfortable and familiar direction. After ASC_CDL adjustments, the inverse matrix converts the adjusted P3_D60 values back to ACES primaries. To summarize, a matrix is applied to ACES to transform ACES rgb to P3_D60 rgb, ASC_CDL is applied in P3_D60 rgb linear (gamma 1.0) light, with adjustments being applied via the natural-feeling P3 red, green, and blue primaries, and then the inverse matrix transforms the pixels back to ACES rgb.

Presentation, System, and Capture Black

With Rec709 on HD-SDI, black is set at 64/1023. However, there is also a common (perhaps unstandardized) use of HD-SDI wherein “full range” is utilized. However, because values in the range 0/1023 to 3/1023 are reserved for sync (due to history, but currently probably unnecessary), the minimum black of full range HD-SDI is 4/1023. However, it is not clear whether 4/1023 means absolute black, or something slightly above absolute black. Note that Rec709 was created with a system dynamic range intended to operate at somewhere around 300:1. Note that current displays and projectors, in the home as well as cinema, usually present a much deeper black than this. The primary parameter was the 2.22 (1/.045) “transfer characteristic”, although this has since been re-interpreted to a 2.4 output-referred gamma as mentioned above. The current use of Rec709 is usually at contrast ratios much higher than 300:1, where it remains adequate, even though 300:1 was the original target of Rec709 parameter optimization.

With ACES, 0.0 means absolute scene-referred black. However, camera lense flare may not allow real values to go to zero except for very dark scenes (or a capped lens). Sometimes a “flare correction” is applied, which can push values down below zero, especially if flare is regional (which is usually the case). Correction for camera and lens flare is inherently complex, and simple flare correction processing is often problematic.

The DCinema specification uses X’Y’Z’ which is CIE 1931 XYZ chromaticity, with a maximum level scaling for setting white (for the “white gamut”), but with a pure 2.6 gamma. This implies that 0/4095 is absolute presentation black.

For the OCES output-referred representation to be device independent, it is not possible to set any single value above 0.0 as a minimum black, but rather OCES is essentially required by its device-independent nature to maintain 0.0 as an absolute presentation black. However, any real device will have flare, and a minimum black that it can achieve, as well as a surround which is likely also somewhere above black (which may reflect a little off of the screen). An output-referred “flare correction” is sometimes used as a catch-all for these issues for a given presentation device, with the affect being that some OCES value above 0.0 is mapped to the minimum device value (such as 0/4095 for DCinema). What this is saying is that the device cannot present absolute black, so an OCES value corresponding to that lifted black should map there for presentation accuracy. However, this is only one possible interpretation. Another is that the dark detail below that value is also of interest, and should be presented rather than being crushed away. This interpretation must lift the absolute OCES black at 0.0 up to the “flared” device minimum (i.e. “best”) black (which for DCinema is 0/4095).

These black representation and processing issues remain a long-standing open topic of discussion.

Negative ACES values

There are several common reasons that ACES pixels might contain negative numbers. For a camera that is calibrated such that a capped-lens black *averages* 0.0, the presence of thermal noise in the sensor requires that values go slightly above and below 0.0. If negative numbers are clipped up to 0.0, then the average will be somewhere between 0.0 and the thermal noise maximum, such as perhaps half way to the maximum.

It is also common to use simple Input Device Transform (IDT) processes such as a matrix transform. However, most camera spectral sensing functions are not linearly transformable into CIE 1931 XYZ. This will yield some colors (usually saturated colors) which have negative values in X, Y, and/or Z, and thus negative values in ACES R, G, and/or B. These IDT negative values can sometimes be quite large, unlike thermal noise near black which is inherently small. There are improved IDT processes being explored and utilized, such as “root-polynomial” transforms which use additional terms to model the color transform to CIE 1931 XYZ. Also, 3-D cross-color lookup tables are sometimes used to define the transform to ACES from the native camera spectral sensing functions.

All such processing, unless the camera sees exactly with CIE-1931-transformable spectral sensing functions (which is quite rare), will result in occasional negative ACES values.

Note also that the IDT is inherently a weighted approximation, attempting to favor the most significant colors (such as face tones). This is perhaps best understood by simple but exaggerated examples. One such example is a camera which senses ultra-violet and/or infra-red light, both of which are invisible to the eye and thus have no weight in CIE 1931 x_bar, y_bar, z_bar color matching functions. For such a pathological camera, infra-red or ultra-violet photons may produce positive pixel values in red, green, and/or blue channels from the camera. Such positive colors clearly are not a correct transform of the invisible light in the scene into visible light in ACES. Another common problem is radar photons causing a periodic washed-out image when photographing aircraft landings.

In addition to these pathological cases, there are also simple common cases of flat spectral sensing function regions where one or two primaries have no R, G, B, distinction difference for a range of wavelengths. Another case is where visible light has no sensitivity in all channels for some spectra region, such as deep red or blue-violet, or even occasionally yellow, green, or cyan for small wavelength regions. Such regions may produce black or may produce incorrect colors (depending on how many channels are involved) for some narrow spectra or combinations of narrow spectra.

Another common way that negative values can be created is when attempting to intentionally crush or perhaps slightly deepen blacks to make them richer. If values are subtracted from the ACES R, G, B values to achieve this look in a “Look Modification Transform” (LMT), then there may be substantial areas of significantly negative numbers in the black regions.

The current RRTs clip negative ACES values as an initial step. Some previous RRT development has explored reproducing negative numbers through the RRT and ODT process. This leaves open what to do with the negative values after the ODT, since they will need to be clipped up to zero corresponding to zero light emitted from the display or projector. The meaning of these negative values is unclear, but the ability to lift negative values into visibility to see into shadows in OCES, or even after the ODT, may be potentially useful for some

applications (to undo an LMT previously applied in ACES which crushed the blacks, for example).

The topic of negative ACES values, and the relationship of these negative values to RRT and/or ODT processing, remains an ongoing discussion. It is clear that there is substantial complexity associated with handling negative numbers in either the RRT and/or the ODT.

For example, a matrix transform (such as ACES to P3) may contain large negative terms (such as -.6 in ACES to P3 in the green-in-red matrix term in the middle of the first row of the matrix). If there is a negative ACES value, then it will be turned into a positive value by the negative term in the matrix row sum. This "folds over" the negative ACES into a potentially highly visible artifact as a positive pixel value, depending on the type of negative value, and the position of the matrix in the processing pipeline. Thus, it has proven useful to allow positive ACES values to be applied to negative matrix terms, and negative ACES values to be applied to positive matrix terms, but to zero out all negative ACES values when applied to negative matrix terms. This then is inherently not-invertible.

This and many other subtle issues (such as the negative portions of the tone curves) make the handling of negative ACES values complex and a topic of ongoing discussion.

Ignoring thermal noise near black, an essential point about one or two negative r, g, and/or b primary values is that they are in some sense potentially valid representations of color, although they have gone outside the gamut boundary or may otherwise be negative due to extreme or unusual cases. Using simplified processes to handle these negative values, such as clipping up to zero, may lead to visible artifacts, noise increase, incorrect and/or unusual colors with unnatural hue and saturation changes, and/or other potentially undesirable outcomes. Some consideration of negative ACES values, and how to best handle them in various possible scenarios, is likely needed.

Negative Values in Rec709

Note that similar issues, some for similar reasons, are present in Rec709. The Rec709 mapping to 10-bits allows Y (luminance) or R, G, and B (in RGB dual-link) to set black at 64/1023, but with values being legal starting at 4/1023. It is quite common, especially in telecine masters, to have large regions of black at 20/1023 or 30/1023, even though these values are undefined in Rec709 in the range between 4/1023 to 64/1023. The digitization of analog video, with inherent cable ringing on horizontal scanlines when crossing vertical edges, was the initial reason given for setting black at 64/1023. However, in practice, the presence of values below 64/1023 remains an issue, especially given that there is no common practice for how to handle these values, nor how to interpret them. Clipping up to 64/1023 is fairly common, but not ubiquitous, and clipping also leaves many issues unaddressed, in addition to inherently not being invertible.

RRT/ODT Implementations and Performance

RRT/ODT development has utilized several different forms of implementation. Initial film-density-based models were implemented in the interpretive "Color Translation Language" (ctl), as were many of the empirical models. The present author translated several of those models into a C SDK, later adding multi-threading. The C implementations included a direct computation, as well as a computation to populate a 3-D Cross-Color Lookup Table (3D LUT), using 101-cubed covering up to ACES value of 64.0. An OpenCL implementation was then explored for the 3D LUT, and later for the full algorithms. Versions were created with interactive

ASC_CDL adjustment, and for HD-SDI 12-bit dual-link output to drive the DLP Projector, the Sony OLED, and the Dolby PRM. The HD-SDI 12-bit dual-link implementations include use of the DVS-Atomix, which can support 4-k via four sets of HD-SDI, as well as the Quadro 5000 HD-SDI.

In order to unify these many SDK modules, the RRT/ODT algorithm was put into an include ".h" file, and translated into macro definitions. The full RRT/ODT algorithm set could then be implemented for all SDK modules, both CPU and GPU, from a single .h include file. This therefore includes the OpenCL version, which is interpreted at runtime. In order to do this, CPU SIMD instructions (such as SSE intrinsic instructions), and OpenCL SIMD instructions (such as 4-wide intrinsic instructions) cannot be used. However, multi-threading on the CPU can be used as well as any compiler-generated SIMD, and OpenCL on both DEVICE_TYPE_CPU and DEVICE_TYPE_GPU implements math functions in parallel (which is not available in SSE SIMD). The OpenCL 3D LUT version runs realtime at 2k resolution on DEVICE_TYPE_GPU using the NVidia Quadro5000 (352 cores), the NVidia 480GTX (480 cores), and the AMD/ATI v7900 SDI (1024 cores) on Linux. It also runs realtime at 2k on DEVICE_TYPE_CPU using the AMD/ATI driver under Linux (on 6Core Nehalem and 4Core SandyBridge processors). It also runs using DEVICE_TYPE_CPU on MacOSX (and may also run realtime on some configurations of CPU's and GPU's on MacOSX, although this is currently untested).

The full-featured directly-computed algorithm (without needing a 3D LUT) runs at about half realtime at 2k on these GPUs, with some simplified RRT/ODT versions, such as a version using a static RRT matrix, running realtime. The Quadro K5000 (1536 cores) will be tested shortly. It is hoped that 2k realtime can be achieved for the full-featured RRT/ODT with direct computation on the GPU, and that 4k realtime can be achieved using the 3D LUT implementation on the GPU.

Note that the DVS Atomix is currently being driven using the "fifo" API using both NVidia and AMD/ATI GPU's. This uses an "up-and-back" RRT/ODT GPU computation. The NVidia "GPU Direct" is not yet implemented in OpenCL, but is anticipated shortly. Note further that the AMD/ATI v7900 SDI implementation using "SDI-Link" will also soon be explored.

Optimizing Future Codecs for Quality

Existing deployed codecs take an approximately common form wherein there is an initial color space conversion (e.g. 4:2:0 or YUV), frequency transform (e.g. DCT or wavelet), quantization, and variable length coding. Motion compensation, using block displacement, is also used in television codecs. The general goal of these codecs is to find low amplitude frequency elements and zero them out, since coding zeros provides much of the compression. Existing deployed codecs, including MPEG-2, H264/AVC, and JPEG-2000 part 1, all use a double-wide deadband around zero. High frequencies are often coded with coarse (intentionally inaccurate) amplitudes, again to reduce coded bits. Color difference channels are often coded with coarse quantization, reduced resolution, or both. The interactions of the different codec processes are difficult to perceive, understand, and optimize. Statistical divergence, leading to a small number of short-duration outliers having larger errors, is common within these deployed codecs.

In addition, the viewing environments wherein the amount of these approximations is optimized represent specific dynamic ranges, color ranges, and surrounds. Specific moving image testing scenes are used. The viewers who judge the moving image quality often vary widely in their ability to discriminate that quality, and their preferences. The bitrates at which these judgments are made and optimized represent specific compression ratio goals.

For future codecs to realize the potential of preserving the image range and quality being created by modern cameras, displays, and projectors, different compression principles are required. Outliers should be bounded, zeros should not be so highly favored, and judgment environments should aim for the highest possible quality (even beyond the discrimination of most viewers). Coding computation should now move from integers to floating point to support the required precision.

The current television practice of sending a single Rec709 format to a wide variety of displays with different adjusted appearances and surrounds, as well as different inherent capabilities, should be replaced by a distribution codec which directly addresses the inherent display differences (and then adjusts, if possible, to a common appearance).

It is herein recommended that new codecs be utilized which can realize the full potential of modern cameras, displays, and projectors by directly coding ACES. It is also recommended that layered coding be utilized, to defer quality level binding until the distribution and display capabilities and the image quality requirements are known. The compression layering can include resolution, quality (signal to noise), and temporal (frame rate) layering. This contrasts with current practice wherein quality limit is "baked into" the signal at encode time.

Further, once fully developed, a common layered compressed master could potentially serve for digital cinema distribution as well as television distribution, unifying the mastered appearance by relying on the ODT process to yield a common appearance for a heterogeneous distribution.

Codec Performance

The present author is demonstrating practical full-range ACES layered coding with high compression ratios, on the order of fifty or one hundred to one, with the RRT/ODT processing being applied to the decoded ACES pixels. Thus, the device-independent coding provides support for numerous display and projector technologies using ODT selection after decoding, including ODT support for high dynamic range and wide gamut devices.

The floating point intra-frame encoder runs in software at about one frame per second for 2k, and about ten seconds per frame for 4k. Decoding is realtime at 2k in software, including running the RRT/ODT interactively using ASC_CDL adjustments in the GPU. Using the latest processors (e.g. dual-8-Core SandyBridge EP) and GPU's (e.g. Quadro K5000), it may even be possible to hit realtime for 4k decoding.

Technology improvements will rapidly expand the range of devices that can implement realtime 2k and 4k ACES decoding, with encoding following shortly thereafter.

Within Reach

Key pieces of the unification being described here are within reach, and are being demonstrated. These include a working ACES system, a working ACES layered floating-point codec, working ODT's for multiple display and projector types, including HDR, high dynamic range cameras capturing a wide gamut at high resolution and high frame rate, ever-increasing and often scalable distribution and storage bandwidth, and other available key technical ingredients.

Summary of Recommendations

Despite our industry's long history of often failing to embrace many valuable system architectural principles, the new digital motion imaging infrastructure that we now have creates an opportunity to consider greatly improving a broad range of system attributes.

The recommendations for consideration being made here can be summarized as follows:

• Begin phasing-out all remaining anisotropic system parameters

• Look to a power-of-two-based horizontal for future television systems if possible, such as 4096 and 8192, instead of 1920-based 3840 and 7680.

• Consider a corresponding power-of-two-based vertical as an alternative to considering 64/27 widescreen, or alternatively propagate the 2048x1080 and 4096x2160 DCinema rasters for future television systems, adding 8192x4320 (for the farther future)

• Look to higher bit-depths in codecs, and begin system-wide enablement and standardization for support of higher dynamic range

• Look to wider gamuts, and begin exploring system-wide enablement of these, while finding ways to reduce gamuts (e.g. within ODT's) back down to the smaller gamuts of existing systems (e.g. Rec709, P3)

• Embrace the ACES concept of scene-referred image processing, with a standardized RRT bridge to a device independent OCES that is display and/or projector referred

• Embrace the ACES concept of ODT's for device-specific conversion of OCES, including high dynamic range devices

• Give thought to negative numbers in ACES

• Standardize the application of ASC_CDL to apply to ACES linear (gamma 1.0) scene-referred light in the LMT position (prior to the RRT). For example, convert ACES data via matrix to P3_D60, then modify by ASC_CDL, and then convert back via the inverse matrix to ACES.

• Continue the ongoing discussion of where to set presentation black, and the meaning and possible practical value of retaining and processing negative numbers

• Begin examining the issue of ambient surround in the home (and elsewhere) and work toward a standard for enabling broad deployment of reference presentation in non-reference ambient surrounds

• Continue exploration of higher frame rates for DCinema

• Find ways to begin to eventually move beyond 50fps and 59.94/60fps television systems (hopefully in a common worldwide system)

• Use temporal layering in deployed codecs to support multiple frame rates.

• Begin exploring the decoupling of pixel clocks using small image update presentation tolerances of a few percent

• Move toward a common layered codec for motion picture exhibition as well as a broad range of television distribution means, and a corresponding common master.

• Move to deployment of a codec which directly codes ACES using high quality layered coding, and embrace the ODT model of device independence after decoding for use with a variety of displays and projectors having a variety of capability levels. Use a common layered codec at low compression ratios (with more layers) for production and contribution, and at high compression ratios (with less layers) for distribution.

Conclusion

Many of the system architecture issues which differentiated digital movie making from digital television are gradually disappearing. The “video look” and “film look” are gradually becoming more similar, especially as higher frame rates are explored for theatrical exhibition. One could say that there is a “pleasing look” that booth are moving toward. ACES is being used in episodic television and in movie production. The inherent technology ingredients in ACES suggest that live television could potentially also move in that direction. High compression ratio coding, using floating point, can provide wide dynamic range preservation while approaching the efficiency of output-range distribution codecs. This can enable extending the dynamic range of motion pictures as well as high definition television and beyond, with current digital cinema cameras already capturing a significantly extended dynamic range. The recognition that the system ingredients for major television and movie system upgrades are within reach might help in furthering the exploration of this potential.

Acknowledgements

Yasuharu Iwaki, Mitsuhiro Uchida, and Alex Forsythe have been the primary authors of both film and empirical models. Alex Forsythe, Scott Dyer, Jim Houston, and Ray Feeney were the primary authors of recent empirical models. Design and implementation input of the ACES system, and the IDT, RRT, and ODT, and help with implementation and support was provided by Josh Pines, Lars Borg, Jack Holm, Paul Chapman, David Ruhoff, Jon McElvain, Walter Gish, Dave Schnuelle, Jim Fancher, Lou Levinson, Andy Maltz, Don Eklund, Thomas True, Will Higgins, and Doug Walker. Many others contributed significantly to the ACES project, which was launched and supported by the Academy of Motion Picture Arts and Sciences’ Science and Technology Council.