Emulating the Film Color Model in Digital Movie Production

Dave Bancroft
Thomson Broadcast & Media Solutions
Reading, United Kingdom

Color image reproduction systems may be designed for pleasing appearance and practical accommodation to the display environment, rather than for "realism." Motion picture films and television broadcasting are examples where the color space representing the reproduction is necessarily different (usually much smaller in its bounded parameter ranges, especially contrast) from the color space representing the original scene.

A complication is that, increasingly, more than one reproduction target is required from the same scene capture; for example, a movie will be shown on network television or released on DVD, as well as being shown on the big cinema screen for which it was originally shot. These different targets involve physically different display devices and also different viewing conditions, notably in regard to dim- versus dark-surround display environments, but there can be differences in viewer expectations, too.

In these cases, therefore, "rendering" is required to convert the light values of the original scene to the much smaller range of values needed to drive the reproduction display device yet maintain the illusion of a large dynamic range. Different renderings are needed to maintain this effect with the different reproduction scenarios.

ISO TC42 has created a document[1] that describes a scheme and structure for the multiple colorimetric image codings involved in the capture ("scene-referred color space"), in the reproduction ("output-referred color space" and in the transformation from one to the other ("rendering"). Although TC42's scope is still photography, the model seems readily applicable to motion picture photography. Appendix A in the same document illustrates a number of examples of color space workflows. One of them[2] is for scanned color negative film and it is reproduced here (Figure 1).

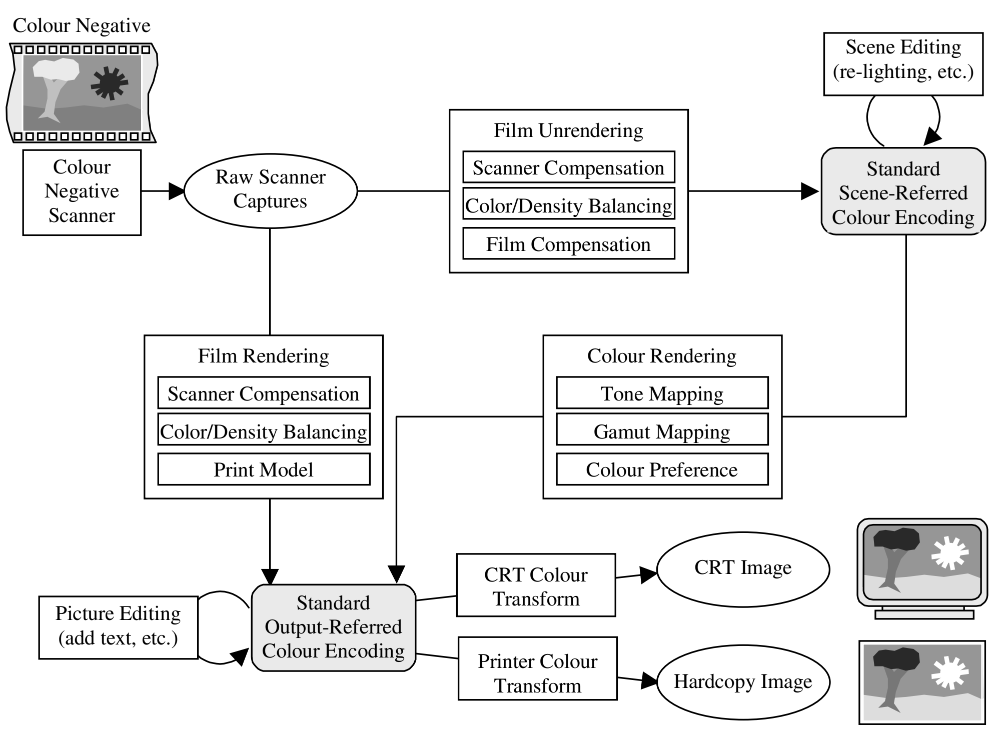

Figure 1. ISO TC42 Generic Workflow for Scanned Color Negative Film (from ISO TC42 22028-1, Figure A.5)

Following the Model with Color Negative Film

When a motion picture film is shot using color negative film in this way, the film stock in the original capture has a very large so-called "exposure latitude." This has evolved to accommodate the creative wish to defer the selection of reproduction contrast range to the post-production stage, rather than deciding it irrevocably—and under difficult conditions—at the time of shooting. In terms of the TC42 document, the range of densities representing the image on the original camera negative is referred to as an "intermediate coding;" strictly, this represents neither scene nor reproduction, because in such a coding, some physical truncation or transformation of the range of colorimetric values has been applied by the capture medium, yet there is still a considerably larger range of values available for use than would be needed or desired in the final reproduction. This is the reason for the term "intermediate coding."

This version of the capture is maintained for quite some way downstream into the creative process; for example, an "interpositive" is a print made from the negative that, despite having a "reversed polarity" in terms of tone range and colors, in fact preserves the gamma and limits of the camera negative and thus its exposure latitude. It is therefore still an "intermediate coding." The creative process that follows will then make artistic judgements on the portion of the captured exposure range to be used in the final reproduction, with the benefit that the latitude available in the "intermediate coding" will ensure that the physical medium's upper and lower limits will not become apparent (unduly crushed shadows or burned-out highlights) in the reproduction.

A side benefit of the continued availability of the "intermediate coding" relatively late in the production process is that more than one "output referred" color space version can be made from the same "scene-referred color space encoding" by selecting from more than one "rendering" algorithm. For example, an encoding can be created for the movie theatre (movie projector color gamut and dark surround viewing conditions), while a different rendering can be applied to the same "intermediate" to create a different version for television broadcast (CRT color phosphor color gamut and dim surround viewing conditions). This is a great advantage in that a single inventory of captured material and a single creative production process (apart from the final color and re-exposure adjustment) yields multiple reproductions for multiple different distribution chains (and revenue streams!).

Based on the above, a significant body of empirical practice facilitating the creation of these different versions has evolved in the movie production business, but only where the original capture has been made on motion picture film.

Electronic Capture

When the capture is made instead with a "television camera," fewer of these desirable attributes have traditionally been available. Dominated by the demands of "live coverage," television production evolved quite separately from motion picture production; for example, it was never envisaged that the captured images could be required anywhere other than on the target viewer's television receiver.

This was acceptable in television broadcasting, because the camera's captured image was always viewable on a high-grade quality control picture monitor, allowing real-time exposure control to be applied quite tightly during capture. Consequently, the opportunity was invariably taken to perform the rendering from "scene-referred" to "output-referred" color encoding in the camera head in real time, as part of the capture process (practically speaking, this was executed in camera head electronic signal processes such as "color balancing," "gamma correction," "scene highlight compression," and "linear matrixing").

The result is that the only output available from a conventional television camera is a "ready-for-use" output-referred version of the content, but which is suitable only for television. If a big screen movie theatre version were then desired, it would be found that various non-linear processing steps applied as part of the output rendering step had irretrievably caused the discarding of those parts of the intermediate coding space that while unnecessary in the television version are now needed for the movie theatre version. The end result would be a movie theatre version that despite the best remedial attempts of directors and colorists[3] bears the unmistakeable imprint of "television": a limited tonal range, notably with "burned-out" highlights among several other unflattering attributes.

For this reason there has been considerable difficulty in getting "digital cinematography" cameras more widely accepted by creative personnel accustomed to motion picture film as their capture medium. There is however a solution but first the requirements will be delineated.

Requirements

1. Modelling in an electronic domain an "extended latitude" scene representation equivalent to film negative and interpositive - a "virtual interpositive";
2. Means for creative control of apparent scene exposure *after* capture by means of extended latitude;
3. Means for "rendering" to multiple "output-referred" versions in post-production/release from single-inventory virtual interpositive:
   - multiple output-referred color gamuts for different display devices
   - multiple tonal rendering intents for different viewing conditions
4. Suitable quality control mechanisms at key stages that preserve positive attributes of photographic film workflow but which add new benefits of electronic real-time interactivity between creators and medium.

Proposed Solution

1. Electronic camera head with NO broadcast-style processing.

2. New capture storage device (no broadcast-style storage device is suitable); new import mechanism into a post-production device normally used only for processing functions on film transferred via telecine; workflow management tools.

Figure 2 compares the two quite different production workflows of the broadcast television medium and the film (movie-making) medium, from scene capture to delivery and display.

The major difference is the point in the flow where "rendering" is performed to transform the colorimetric values from the original scene to a suitable set of values for display. In the broadcast television workflow, because the image is visible and because many broadcasts must be transmitted live, this rendering is performed immediately in the capture device (television camera), with the accuracy verified interactively on a video monitor. The output from the camera is therefore "ready for display" on well-standardized consumer devices.

In the photographic film workflow, the image is not visible at the time of capture, so instead, careful exposure measurement is made by the Director of Photography (D.P.) before shooting and this approach is assisted via a chemical formulation in the camera negative film stock that provides a very wide exposure latitude. This reduces any loss of the extremes of the eventually-desired tonal range through under or over-exposure. At the same time, the camera film stock makes no attempt to synthesize color-matching values for the eventual display; that is done subsequently in the printing stages. The negative film stock is therefore in effect an intermediate "carrier" of predominantly scene-referred image encoding that allows the rendering to an "output-referred" encoding to be deferred to a later stage than the capture stage. That particular output target is singularly the cinema projector, which is well specified[4] in terms of luminance, color temperature and other parameters, as well as in terms of recommended ambient viewing conditions (no distinction is made here between conventional and electronic cinema projectors, since the latter are being designed to replicate the color characteristics of film projectors as closely as possible).

Figure 3 shows the concept of synthesizing a new workflow from elements of both television and film approaches. Such a workflow is applicable to making movies (but not to live broadcasts). The medium is an electronic, not chemical, image representation, and therefore the capture device bears a strong resemblance to a "television" camera, but every attempt is made to preserve the "raw" values generated by the camera's sensors (usually solid-state, e.g. CCDs) and deliberately *not* to transform them via any rendering algorithms, especially not non-linear algorithms.

In post-production, the transformation is similar to film, but the opportunity is taken to benefit from being able to perform multiple renderings to multiple output devices and viewing conditions from the same captured material.

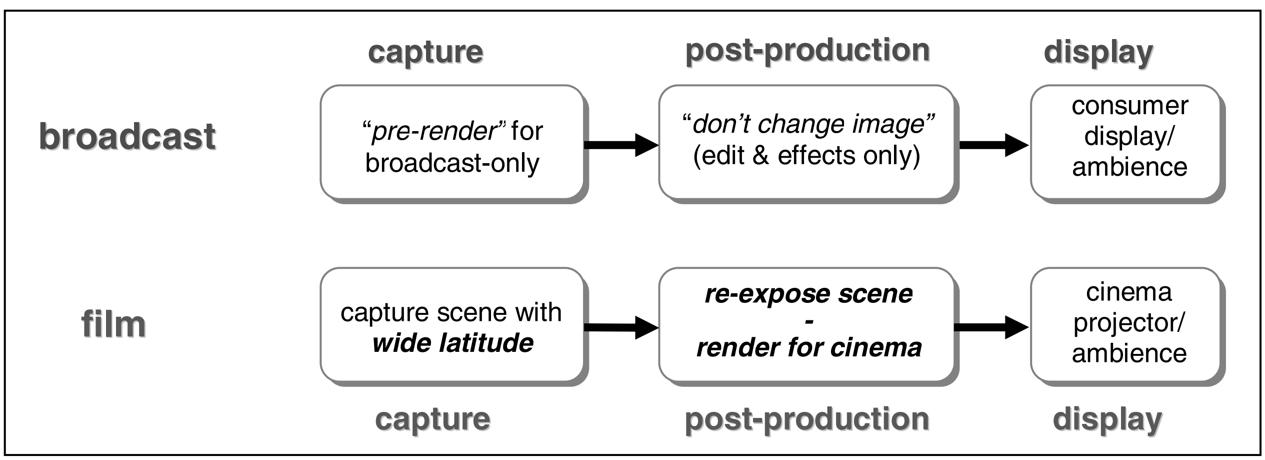

Figure 2. Comparison between broadcast television and film production workflows

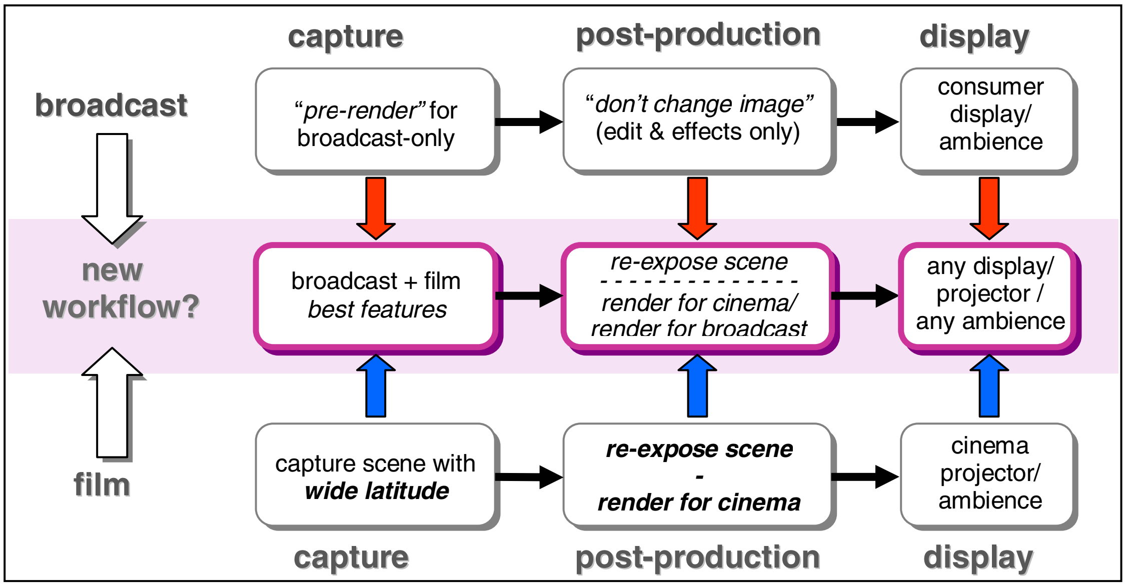

Figure 3. Synthesis of "New Workflow"

Practical Execution

Figure 4. Digital Cinematography Camera with "Intermediate Coding"

Figure 4 shows a digital cinematography camera[5] utilizing these principles and Figure 5 shows it in combination with a field recording device. This particular camera uniquely outputs "raw" values from its linear CCD sensors (after analog-to-digital (A/D) conversion). These values extend from the sensor noise floor to pixel photosite saturation and at this stage of the flow are uncalibrated or balanced to any color space's range of values. Because the camera's output thereby emulates the function of a processed color negative (but one which has been inverted into an interpositive), the output is also referred to as a "virtual interpositive." It should also be noted that preserving "intermediate coding" values requires more data capacity than a typical broadcast "output-referred" coding. Consequently the associated field recording device uses a transportable hard disk system (in the central "pod" in Fig. 5)), capable of preserving the camera's signals in an "RGB" form[6]; broadcast video tape recorders, in contrast, can only store the component set of luma[7] plus sub-sampled color difference signals prevalent in broadcasting, so are unsuitable for this application.

Figure 5. Camera with Disk-based Field Recorder

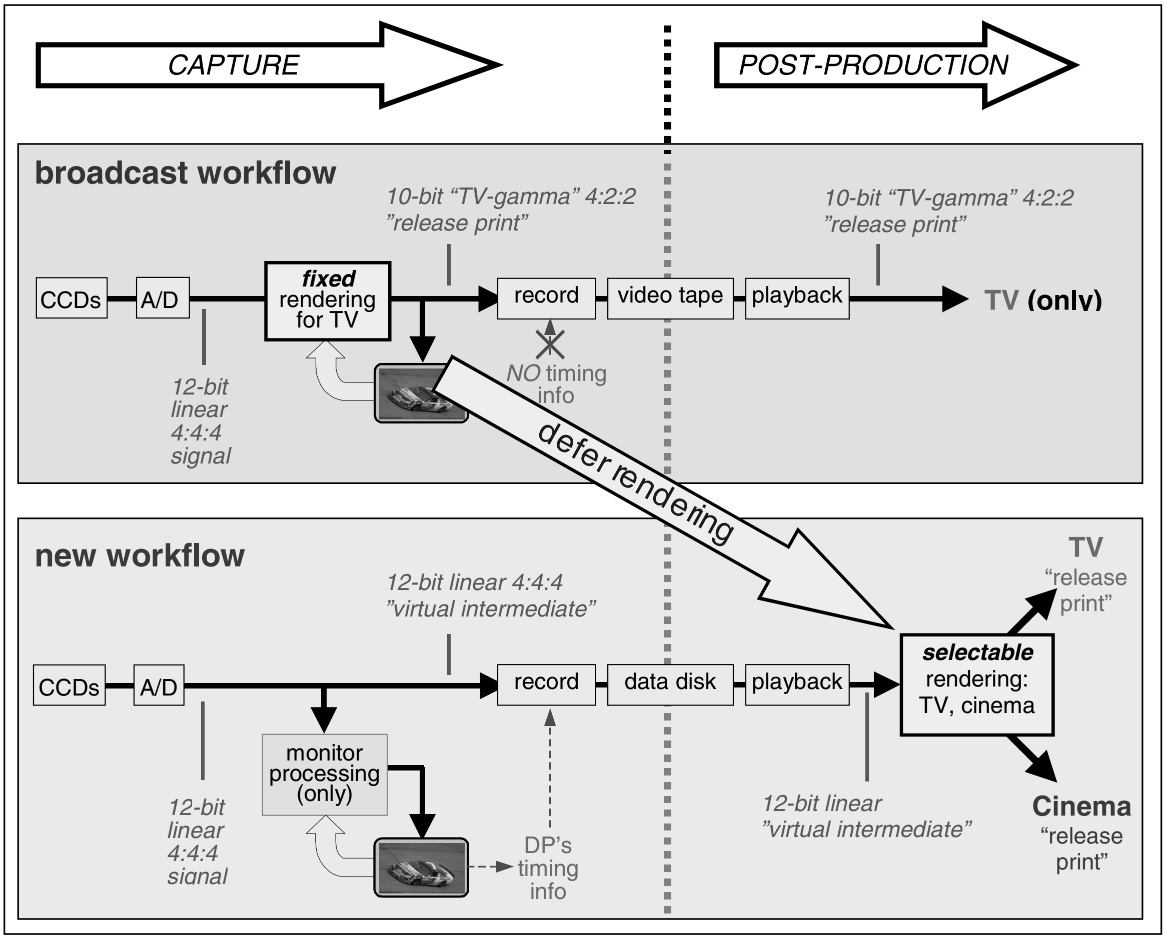

Figure 6. “New Workflow” - with “Deferred Selectable Rendering”

Figure 6 shows how the “new workflow” evolves from the traditional broadcast television workflow by means of these two critical components. In the upper part of the figure, the block shown as “fixed rendering for TV” consists of linear operations such as color balancing and linear matrixing, and non-linear operations such as “TV gamma correction” (pre-correction for the CRT transfer characteristic) and scene highlight compression. The diagrammatic representation of a picture monitor with a feedback path into the rendering block refers to the realtime, interactive quality control applied both in the technical and creative sense at this stage. Film directors would call this “timing,” but it is noteworthy that none of the values of the television “timing” adjustments are preserved with the recorded image; in post-production it is therefore not possible to know what they were in capture. The resulting “rendering” is “fixed” in the sense that it is aimed exclusively at the particular characteristics of the color CRT display in the consumer’s home, with associated “dim surround” viewing conditions.

In the lower part of the figure, this rendering operation has now been moved to post-production. Exactly the same processing operations are made available as before, but multiple sets can be used to make differently rendered versions. The diagram depicts renderings for broadcast television and for cinema, but other versions are possible; for example, it would be possible to make a “home cinema” version on DVD, intended for viewing conditions somewhere between “dark surround” and “dim surround.” The rendering has therefore been changed from fixed to selectable and this has been achieved by deferring its point of application.

A comment on the image monitoring in the field capture stage is appropriate. If the new model were simply emulating film, no monitor would be used at all. But it would be a lost opportunity not to exploit the instant image that is available; it is invaluable for shot setup, confirmation of successful capture and instant distribution for editorial and other purposes. However, the critical distinction is that the monitor is *not* used to guide the D.P. into adjusting any processing in the main signal path. Instead, adjustments are made in the "monitor processing" (which mimics the rendering that will occur later in post production) until an approximately correct image is obtained. The values of these adjustments then represent the "DP's timing information;" this is recorded along with, but separate from the main image signal on the image recording device. This information is then recovered as metadata in post-production and used as a guide to what kind of exposures the DP was trying to achieve at the time of shooting; these can be much more closely achieved (and matched to those of other shots) in the more regulated conditions of post-production.

It might be thought that a "broadcast" version of the content could simply be "re-rendered" to create a cinema version. This would at least allow an unmodified broadcast television camera with its associated monitoring facilities to be used for "digital movie making." Unfortunately, some of the operations performed in the camera head processing, such as clipping of negative color values used in linear matrixing prior to applying the signal to the output interface (which accepts only positive values), result in the permanent loss of information that would be needed when re-rendering to the different color gamut of a cinema projector. The re-rendering is therefore inadequate, the results comparing unfavourably with those obtained from a film original.

An analogy with the terminology of the film world may clarify this point: a conventional film camera does not produce a "release print:" it produces an exposed, then processed, negative, with the latitude characteristics described earlier. For convenience (and protection of the original) this is quickly copied to produce an "interpositive." Since the interpositive has a transfer characteristic power law ("film gamma") of unity, it equals the negative in terms of density range, hence exposure latitude and color separation characteristics; thus it acts equivalently to the negative as input to the "scene-referred-to-output-referred rendering" stage. Both negative and interpositive are therefore generically known as "intermediates." The object of the electronic model for the new workflow is therefore that the capture stage should stop producing "release prints," and should instead produce an "intermediate," but it is now an electronic intermediate, with the same or nearly the same exposure latitude and color gamut potential as the film intermediate.

Color Model Viewpoint of Process

The ISO TC42 22028-1 model was referred to earlier. Figure 7 shows how the practical implementation described above could be expressed in terms of that model. This is something of a hybrid illustration as it combines both theoretical and practical elements. It shows, however, what needs to be done to add digital cinematography to an existing color management workflow for multi-media output. The key points are:

- the use of a conventional film camera and negative film process (sub-flow (a)) achieves the intent of the ISO TC42 model, allowing "cene editing" to be performed
- substituting a broadcast television camera with broadcast display processing embedded in the camera (sub-flow (c)) is not an acceptable film capture replacement because it effectively prevents scene editing
- substituting a digital cinematography camera that eschews front-end output-referred (display) pre-processing (subflow (b)) is, however, acceptable because it allows scene editing with similar freedoms to the film camera sub-flow
- by this means choice of capture medium is achieved
- respecting the remaining provisions of the TC42 model also allows freedom in distribution medium to be achieved, yet without requiring multiple-color-version content inventory.

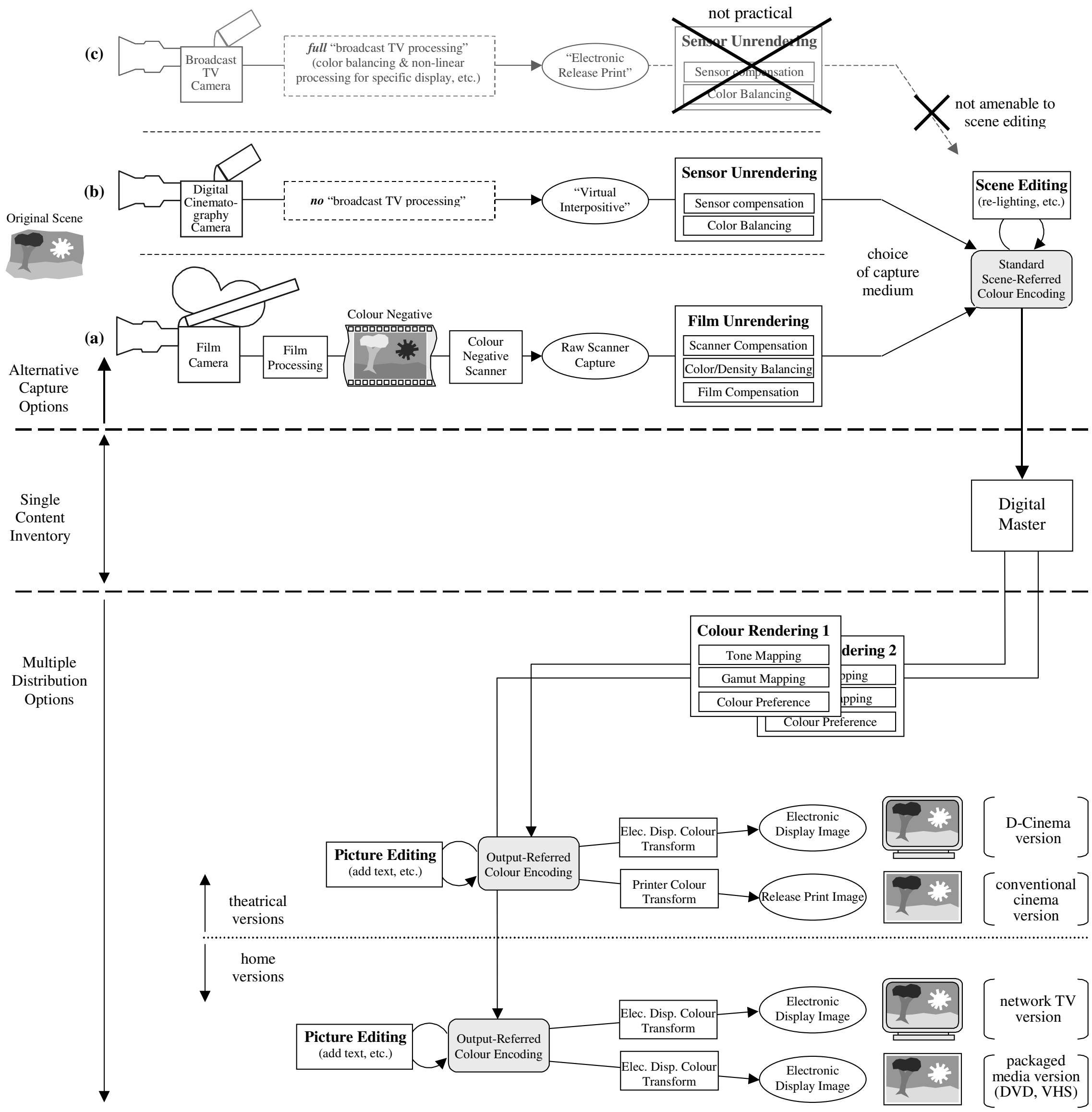

Figure 7. ISO 22028-1 Schema Expanded to Multiple Source Media Application

Conclusion

The discipline of using specified image states in a color space and color encoding model of a complete end-to-end image reproduction system has distinct benefits and acknowledgment is made to the ISO TC42 committee's work in this regard. In the motion-picture production example described, it was desired to obtain image capture from digital camera sources in seamless juxtaposition with that from conventional film cameras, and to obtain multiple renderings to widely differing output devices and viewing conditions in both cases. Modeling the production chain in this way provides enhanced clarity as to the correct positioning of key processing stages in the chain. New devices for practical execution of the concept have become available, in the form of a digital cinematography camera and a field recorder, both eschewing traditional broadcast video signal representation, storage and transfer techniques in order to achieve the new objectives.

Annex 1: Specifications of Output Reproduction Targets

A.1 The Television Display and Viewing Environment

The colorimetry of the television display is defined in several places. The definition is possibly at its most precise in the specifications for high-definition television, since this was defined more recently than standard-definition television, thus exploiting the availability of more sophisticated processing in both origination equipment and in consumer television displays. An appropriate source is ITU-R Recommendation BT.709-4, "Parameter Values for HDTV Standards for Production and International Programme Exchange. The relevant section of the latter is "Opto-electronic Conversion" and for convenience, it is reproduced here (this then needs to be qualified by the sRGB standard to define the target monitor more clearly):

1 Opto-electronic conversion

Item	Parameter	System Values									
		60/P	30/P	30/PsF	60/I	50/P	25/P	25/PsF	50/I	24/P	24/PsF
1.1	Opto-electronic transfer characteristics before non-linear pre-correction	Assumed linear									
1.2	Overall opto-electronic transfer characteristics at source(1)	$V = 1.099\, L^{0.45} - 0.099$ for $1 \geq L \geq 0.018$ $V = 4.500\, L$ for $0.018 > L \geq 0$ where: L: luminance of the image $0 \leq L \leq 1$ V: corresponding electrical *signal*									
1.3	Chromaticity coordinates (CIE, 1931)	x					y				
	Primary – Red (R) – Green (G) – Blue (B)	0.640 0.300 0.150					0.330 0.600 0.060				
1.4	Assumed chromaticity for equal primary signals (Reference white)	D_{65}									
		x					y				
	$E_R = E_G = E_B$	0.3127					0.3290				

A.2 The Cinema Projector and Viewing Environment

This is defined by ANSI/SMPTE Standard 196M-1995 (www.smpte.org)

A.3 Home Theatre

Home theatre is essentially a closed circuit TV system based on the playback of packaged media, notably DVD and display on large screens, usually "indirect view," i.e. front- or rear-projectors. The colorimetry is not standardised; however, at the present time, the displays, although using a wide variety of physical light source and modulation methods, are currently all designed (with color gamut mapping mechanisms at their inputs where necessary) to accept "television" signals that follow the standards referred to in **A.1**; in other words, colorimetrically they aim to emulate the color-phosphor CRT. The distinction with reference to standard television is in the color balancing preparation performed in post-production. A different assumption is made about the viewing environment, i.e. that the room illumination level is somewhere in between that of the cinema (very dark surround) and conventional television (so-called dim surround). This affects the rendering intent, which needs to be somewhere between the approx. 1.2 system gamma of television and the 1.5 of conventional cinema. In addition the coefficient values used in the linear matrix that is employed to reduce saturation loss in the display in colors close to the spectral locus tend to be optimized differently compared to television.

Annex 2: Evaluation Methods

Based on the empirical nature of long-established operational practices in the movie industry when using conventional film, no acceptance of an electronic substitute for the conventional film workflow will occur until at least some subjective tests have been passed. These will inevitably involve side-by-side comparisons between film and electronic reproduction of the same source scene. The "Macbeth" color chip chart is commonly employed in conjunction with a series of camera exposures of different value. It is common to run such tests with conventional film from two stops under-exposure through nominal to two stops over-exposure, although extension to plus and minus four stops is occurring more and more now. The intention is to see where the film stock or electronic sensor exhibits a clear onset of "overload" or "crushing" of the tonal curve at either the black or white extremities and to note any difference in these thresholds between the two capture medium types. Since three color channels are involved in both media types, such "overload" takes the form not only of luminance overload but also of differential effects leading to hue shifts and saturation changes.

References and Endnotes

1. ISO TC42 22028-1, Photography and graphic technology - Extended colour encodings for digital image storage, manipulation and interchange - Part 1: Architecture and requirements.
2. ibid, Annex A, Figure A.5.
3. The technicians who alter tone scale and color balance in a reproduced image for both remedial and creative purposes.
4. Society of Motion Picture and Television Engineers (SMPTE), standard ANSI/SMPTE 196M-1995, "Indoor Theater and Review Room Projection -Screen Luminance and Viewing Conditions.
5. Viper FilmStream camera (Thomson Broadcast & Media Solutions, Breda, The Netherlands).
6. Example shows "Director's Friend df-cineFS" recorder, with "XD Reel" disk module (Directors Friend GmbH, Cologne, Germany).
7. "Luma" is a term now used to refer to the signal component indirectly representing luminance in a broadcast television signal; this signal is in fact matrixed from non-linear transformed RGB signals, yet was previously incorrectly referred to by the linear-light term luminance. For a full explanation, see Poynton, C., "Digital Video and HDTV - Algorithms and Interfaces," Morgan Kaufmann 2003.