Perception of color and image quality of films projected in cinema or displayed on TV screen

Alain Trémeau

Lab. LIGIV – Univ. Saint Etienne – France
http://www.ligiv.org

CREATE - Managing Colour in Digital Processes & the Arts
Bristol, 18th - 20th September 2007
Outlines

1. Image quality assessment based on technical factors
2. Image quality assessment based on visual perception factors
3. Color management in the digital chain
4. Film look creation
5. Film look validation
Image rendering varies significantly from one display technology to another one.

The first work to do before analyzing the quality of a displayed image is to analyze the display device used.
Display technology assessment factors

<table>
<thead>
<tr>
<th>Spatial Uniformity (in luminance and chrominance)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gray scale dark artifacts</td>
</tr>
<tr>
<td>Display aging</td>
</tr>
<tr>
<td>Ambient light</td>
</tr>
<tr>
<td>Viewing angle</td>
</tr>
<tr>
<td>Screen deep (Sharpness) of front projection (direct view) or rear projection</td>
</tr>
<tr>
<td>Screen size (Screen door effect) of front projection (direct view) or rear projection</td>
</tr>
<tr>
<td>Presence</td>
</tr>
<tr>
<td>Consumption</td>
</tr>
</tbody>
</table>
Display technology assessment factors

The three-component reflection model

*Specular, Lambertian and Haze effects due to viewing angle.*
Display technology assessment factors

| Grey-Scale and color | - Black level  
|                     | - Peak Intensity white  
|                     | - Contrast ratio  
|                     | - Checkerboard display contrast  
|                     | - Color primaries (color gamut)  
|                     | - White point  
|                     | - Gamma curve  
|                     | - Transfer function: brightness (luminance) versus signal intensity  
|                     | - Brightness scale  

| Brightness ratio has an influence on: | - Image contrast,  
|                                       | - Image brightness,  
|                                       | - Hue,  
|                                       | - Color saturation.  

Grey scale and color accuracy factors
Display technology assessment factors

Luminance of dark state (left), bright state (center) and the resulting luminance contrast ratio (right) in function of the viewing angle, in a polar coordinate system.
Grey Scale and color Transfer functions

Transfer function: brightness (luminance) versus signal intensity.
Gamma correction

The RGB values of projector devices (e.g. LCD projectors) are not in proportion to the spectral power distribution of the input signal.

The main difficulty a user has to face is to select the best linearization method (e.g. gamma-correction) which minimizes the fitting errors for all three channels.

The first step is to characterize a projector device in determining its spectral sensitivities, in order:

1. to fit sensitivity functions to set of color matching functions,
2. to compute fitting error (ΔE) corresponding to the deviation of these functions from the color matching functions.
Grey scale and color accuracy

The color gamut of a display is defined by the area inside the triangle formed by connected its primary colors.

The accuracy of color reproduction is generally what matters the most.

Primary colors of several display technologies (CRT, LCD, DLP, Plasma).
Grey scale and color accuracy

The ITU-R BT.709 is actually the official standard for High Definition television.

The different display technologies (standards) produce different color renderings of any image.

The differences are most apparent in images that include highly saturated colors.

Color gamut of official standards (NTSC, ITU-Rec709 and EBU) and color gamut of the Visual Human System.
Grey scale and color accuracy

Saturation of primary colors of several display technologies (CRT, LCD, DLP, Plasma).

Primary and white point variation due to a change of lighting source (6500K->9300K).
Example of reproduction of wrong colors.

Grey scale and color accuracy

Mobile LCD displays have a rather dim backlight and thin color filters, resulting in poor brightness and color reproduction.
Subjective assessment factors

Parameters affecting the perception of displayed images

The image rendering depends also of the visual attention of the observer who watches the displayed images.
Subjective assessment factors

The color of the surround (varies with the use of a backlight) has an influence on the color perception, especially on the color contrast of the displayed image.

Technology Ambilight (© Philips).

CREATE - Managing Colour in Digital Processes & the Arts
Bristol, 18th - 20th September 2007
The main problem an user has to face is how to control the relationships between display parameters and how to take into account viewing conditions and image content.

Example of color corrections an user can apply to increase the image quality.

CREATE - Managing Colour in Digital Processes & the Arts  
Bristol, 18th - 20th September 2007
Color management in the digital chain

How characterizing and ensuring the conditions of display, even if display media change?

Film projector (35 mm) is here the reference, meanwhile the digital projector image has to be color-managed.

Digital film reproduction chains
Color management in the digital chain

The first step is to characterize a projector device in determining its spectral sensitivities, in order:

1. to fit sensitivity functions to set of color matching functions,
2. to compute fitting error ($\Delta E$) corresponding to the deviation of these functions from the color matching functions.
Calibrated Digital Intermediates (DI) workflow.

- The target device is a digital projector (at bottom left side). The projector is set to film look using a Look Up Table (LUT) stored in a color box and applied to the RGB signal.

- The LUT is based on a forward device model for film and an inverse device model for the projector: - the forward device model describes mathematically the film printing and projection chain; - the inverse device model describes the digital projection process.
Color management in the digital chain

The device models are established by device characterization.

- *Device characterization is the process of building a device model from measurements.*

For a display device, measurements are the set of RGB input values and measured XYZ values of displayed colors.

- *RGB are device dependent color input signals of the considered device,*
- *XYZ are values from CIE 1931 XYZ space for 2-degree observer.*

Device characterization generates a device model offering the following functionalities:

- *Interpolation of XYZ output values from given RGB input values (forward model)*
  \[(X, Y, Z)^T = f_{XYZ}(R, G, B)\quad \text{where} \quad (\cdot)^T \text{ is transpose operation.}\]

- *Interpolation of RGB input values from given XYZ output values (backward model)*
  \[(R, G, B)^T = f_{RGB}(X, Y, Z)\]
Color Management in the digital chain

Color perception relies to color constancy (chromatic adaptation) phenomenon which allows the Human Visual System to adapt to widely different viewing environments.

The second step consists in using a color space transform (including a gamut mapping transform) which modifies the rendering of the output image (e.g. a projected image), proportionally to the input image (an image data file, a film), according to different factors.
In order to take into account viewing conditions and visual attention, color management processes include a color appearance model.
Color Management in the digital chain

These factors depend firstly of viewing conditions:

• brightness/contrast settings,
• white point (illumination condition),
• white balance (max-white to min-black),
• color balance (gamma-curves for R, G, B-channels),
• white balance versus color balance,
• white scale and color scale,
• color gamut.
• luminance and chromaticity of the darkened screening room (mesopic condition).
Color Management Method

Knowing the lighting conditions and the content of the scene (e.g., the color of walls), it's possible (by using a color constancy model) to compute the ambient illumination under which the scene was viewed.

All images of this video can be colorimetrically corrected by calibrating their color distribution to those of the Macbeth color chart.

Example of video acquired under **illuminant C**

The same scene viewed under **illuminant D65**

From one lighting condition to another one, the colors change significantly.
Color Management Method

These factors depend firstly of viewing conditions:

- brightness/contrast settings,
- white point (illumination condition),
- white balance (max-white to min-black),
- color balance (gamma-curves for R, G, B-channels),
- white balance versus color balance,
- white scale and color scale,
- color gamut.
- luminance and chromaticity of the darkened screening room (mesopic condition).

Lastly, they depend of spatial parameters:

- homogeneity of luminance and color (light distribution, chroma distribution),
- contrast and sharpness (MTF),
- image content (number of Regions of Interest, their size and their disposal).

The last step consists in analyzing the effects due to relationships which exist between these parameters.
Device characterization is not the focus of this paper. In general, a proprietary solution is used.

Characterization errors can be \textit{one source for errors in digital “film look”}. They may be caused by:

- measurement errors (noise),
- dynamic flare (mutual reflection between screen and theater surfaces),
- imaging chain errors (spatial distortions, temporal variations)
- inherent characterization errors (model errors, gamut mapping).

In the following slides we will speak about “film look” creation, i.e. about the \textit{calibration} of a target device in order to have a “film look” reproduction.
Film look creation

The target device is set to “film look” using Look Up Tables (LUTs).

RGB-RGB LUTs are built from:

- a forward film chain device model (RGB to XYZ) and
- an inverse digital chain device model (XYZ to RGB).
Film look creation

The target device is set to film look using Look Up Tables (LUTs).

RGB-RGB LUTs are built from:
- a forward film chain device model (RGB to XYZ) and
- an inverse digital chain device model (XYZ to RGB).

A LUT can be represented as:

\[
\left\{ \left( R_j, G_j, B_j, \hat{R}_j, \hat{G}_j, \hat{B}_j \right) \mid \forall j \in [0, P^3] \right\}
\]

with \( P \) being the resolution of the LUT, e.g. \( P=64 \).
Film look creation

A LUT for calibration of the digital chain in “film look” is calculated as follows:

\[
(R_j, G_j, B_j) = \left( \frac{p(2^M - 1)}{P - 1}, \frac{q(2^M - 1)}{P - 1}, \frac{j(2^M - 1)}{P - 1} \right)
\]

\[
(\hat{R}_j, \hat{G}_j, \hat{B}_j) = f^{TARGET}_{RGB} \left( f^{REF}_{XYZ} (R_j, G_j, B_j) \right)
\]

with \( M \) : is the bit depth of each color feature, e.g. \( M = 10 \) bits.

\[
p = j \mod P^2; \quad q = j \mod P; \quad 0 \leq j < P^3
\]

where \( f^{TARGET}_{RGB} \) is the inverse device characterization model of the target device and \( f^{REF}_{XYZ} \) is the forward device characterization model of the reference device.
Errors in the generation of the “film look” LUT can be a second source for errors in digital “film look”.

They may be caused by characterization errors and inherent linearization or quantization errors of the LUT.
Film look validation

Three subjective tests can be used for validation of “film look” reproduction on digital displays in a DI postproduction workflow.

The main goal in DI postproduction is to evaluate the precision of color reproduction on a target display (a digital projector) calibrated to “film look”.
Film look validation

Subjective tests are then applied to images displayed:

• by the film chain (digital images printed to film and then projected) and
• by the digital chain (images projected by a digital projector) set to “film look”.

• The Film chain consists of film recording (printing a master negative from a Digital Intermediate) and film printing (copying a positive from the master negative).
• The digital chain contains a color box that applies a color transform to the RGB signal such that the target display is set to film look.
• The film look is generated by a 3D LUT that is calculated from a film chain device model and a digital chain device model such that the digital chain reproduces “film look”.

Additionally to subjective tests, “film look” can also be validated by objective measurements, for example using colored patches.
Validation and subjective evaluation of “film look” on digital display devices.
Film look validation

• Subjective psycho-visual test methods require human viewers, expert or non-expert, to rate the quality or difference in quality of two clips.

• Unfortunately, humans are biased when comparing good “traditional” film with digital. Subjective tests without bias are difficult to realize.

• In most testing scenarios two clips are considered, these two clips differ in the fact that one will be the reference and the other will be processed in some manner.

  • Subjective assessment are costly and time-consuming.

  • This type of assessment is particularly necessary in critical situations such as final product evaluation and standardization processes where quality must be assured.

    • Subjective assessment methods have been used reliably in the past to evaluate video image quality but not to evaluate digital film quality.
Film look validation

Three tests may be used, they have the following characteristics:

• **Use of real film content.** To enable final assessment comparable to real application case, short segments of real film content with high importance to color quality is used for the tests.

• **Selection of content according to use cases.** The segments of test content are selected corresponding to specific use cases such as “hue of skin color”, “saturation of blue sky” and “tone of night scene” in order to cover all important regions of film gamut and allow for technical validation.

• **Use of a 12 persons test group.** As compromise between statistical evaluation and limited availability of specialized test persons, the size of test group is limited to 12 persons.

• **Restriction to US market.** For each evaluation, content selection, criteria and experts are chosen according to ethnical and cultural habits of Northern America.
Film look validation

Example of video for which the “saturation of blue sky” is an important parameter of study.
Film look validation

Example of video (of very bad quality) for which the “hue of skin color” is an important parameter of study.
Film look validation

Three tests are used:

- **Test 1** is designed to give a result of high confidence since observers are not biased (introduction of 2 unknown digital sources).
- **Test 2** has a proven bias (film and digital sources are easily recognizable) but is close to the real application case of color correction.
- **Test 3** has been added to collect natural language type comments of “golden eyes”.
Test no. 1: moving picture test

This first test:

- firstly projects the film reference version,
- next projects two digital film look versions, one after the other,
  (B can be a film look reference and A can be a new digital version to test);
- asks for quality assessment according to a quality scale.

This method is called the Double-Stimulus Continuous Relative Quality Scale method (DSCRQS).
Test no. 1: moving picture test

Example of bright image for which the “saturation of fruits color” is an important parameter of study.
Test no. 1: moving picture test

Example of dark image for which the “contrast of dark colors” is an important parameter of study.
This figure shows the basic test cell.

- A test cell begins with a test logo to indicate the start of a new test cell.

- A 10s segment of test content is first projected as film two times and interlaced with 3s medium gray sequences.

- Then projection is changed to digital media and the same segment of content is projected iteratively during 5s for A and B interlaced with 3s medium gray sequences and followed by 30s medium grey for voting.
Test no. 1: moving picture test

Comparison of two stimulus (A and B) to film.

Since viewers are biased when watching and recognizing the film reference version of test content, tests include two digital stimuli after showing the film stimulus while assessments are done on the two digital versions with respect to film.
Test no. 1: moving picture test

This first test:

- firstly projects the film reference version,
- next projects two digital film look versions, one after the other,
  (B can be a film look reference and A can a new digital version to test);
- asks for quality assessment according to a quality scale.

This method is called the Double-Stimulus Continuous Relative Quality Scale method (DSCRQS).

The test persons are asked to note films A and B quality on a grading scale according to specific questions on hue, saturation and contrast.
Quality grading scale for test no. 1
Test no. 1: moving picture test

This first test:

• firstly projects the film reference version,
• next projects two digital film look versions, one after the other,
  (B can be a film look reference and A can a new digital version to test);
• asks for quality assessment according to a quality scale.

This method is called the Double-Stimulus Continuous Relative Quality Scale method (DSCRQS).

The test persons are asked to note films A and B quality on a grading scale according to specific questions on hue, saturation and contrast.

The viewers are allowed to assess during the voting period only. A test is run with 6 persons at once according to a test set-up.
Viewing room set-up for subjective test no. 1
Test no. 2: still picture butterfly test

This second test:

• projects side-by-side two versions of a still picture
  (e.g. film and digital film look);
• asks for artifact assessment according to an impairment scale.

This test is called the Butterfly Test Method (BT).
Test no. 2: still picture butterfly test

Step 1: Selection of the still picture according to use cases. The test pictures are selected corresponding to specific use cases such as “hue of skin color”, “saturation of blue sky” and “tone of night scene” in order to cover all important regions of film gamut and allow for technical validation.
This figure shows the basic test cell.

• A test cell begins with a test logo to indicate the start of a new test cell.

• A 15s medium grey period for adaptation is followed by the side-by-side presentation of the reference (right side: film) and A (left side: digital film look).

• At the end, 5s of medium grey are displayed for voting.
Test no. 2: still picture butterfly test

This second test:

- projects side-by-side two versions of a still picture (e.g. film and digital film look);
- asks for artifact assessment according to an impairment scale.

This test is called the Butterfly Test Method (BT)

The test persons are asked to assess the impairment of A with respect to the film reference on a grading scale according to specific questions on hue, saturation and contrast.
<table>
<thead>
<tr>
<th></th>
<th>Impairment scale for subjective test no. 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>The same</td>
</tr>
<tr>
<td>+1</td>
<td>Slightly different</td>
</tr>
<tr>
<td>+2</td>
<td>Different</td>
</tr>
<tr>
<td>+3</td>
<td>Much different</td>
</tr>
</tbody>
</table>
Test no. 2: still picture butterfly test

This second test:

- projects side-by-side two versions of a still picture (e.g. film and digital film look);
- asks for artifact assessment according to an impairment scale.

This test is called the Butterfly Test Method (BT)

The test persons are asked to assess the impairment of A with respect to the film reference on a grading scale according to specific questions on hue, saturation and contrast.

The viewers are allowed to assess during the voting period only. A test is run with 3 persons according to a test set-up.
Test no. 2: still picture butterfly test
Viewing room set-up for subjective tests no. 2
Test no. 3: still picture in-depth butterfly test

This third test:

- projects side-by-side two versions of a still picture (e.g. film and digital film look);
- asks for in-depth assessment with unlimited time and free comments.

This test is called the in-depth Butterfly Test Method (IDBT).

Test 3 is very similar to test 2 except that presentation and voting time are unlimited and comments are in free text. This test responds to the fact that test persons are experts.
This figure shows the basic test cell.

- A test cell begins immediately with the image under test, the observers adapt to the image.
- The duration is unlimited, the presentation is side-by-side of the reference (right side: film) and A (left side: digital film look).
- Presentation and assessment takes place at the same time.
Test no. 3: still picture in-depth butterfly test

This third test:

• projects side-by-side two versions of a still picture (e.g. film and digital film look);
• asks for in-depth assessment with unlimited time and free comments.

This test is called the in-depth Butterfly Test Method (IDBT).

The test persons are asked to note comments in free text. A single test is run with 3 persons according to the same test set-up as for test 2.
Viewing room set-up for subjective tests no. 3 (the same as test no. 2).
Conclusion

• The motivation of our recent researches is to develop a software package to test and to compare dynamically different cinema processing and post-processing algorithms.

• The ultimate goal of these researches is to implement, at long range, an open and unified color imaging system (i.e. a "color toolbox") to set up a favorable context for the evaluation and analysis of color imaging processes.

• This supposes, as a minimal pre-request, to develop a video processing tools demonstrator and a standard test video database to validate post-processing algorithms such as color enhancement or color correction processes.
References


- ASC/DCI StEM mini-movie, http://WWW.dcimovies.com
References


Questions ?