Evaluation of color grading impact in restoration process of archive films

Karel Fliegel[a], Stanislav Vítek[a], Petr Páta[a], Petr Janout[a],
Jiří Myslík[b], Josef Pecák[b], and Marek Jícha[b]

[a]Czech Technical University in Prague, Technická 2, 166 27 Prague 6, Czech Republic;
[b]Film and TV School of Academy of Performing Arts in Prague, Smetanovo nábřeží 2,
116 65 Prague 1, Czech Republic.

ABSTRACT

Color grading of archive films is a very particular task in the process of their restoration. The ultimate goal of color grading here is to achieve the same look of the movie as intended at the time of its first presentation. The role of the expert restorer, expert group and a digital colorist in this complicated process is to find the optimal settings of the digital color grading system so that the resulting image look is as close as possible to the estimate of the original reference release print adjusted by the expert group of cinematographers. A methodology for subjective assessment of perceived differences between the outcomes of color grading is introduced, and results of a subjective study are presented. Techniques for objective assessment of perceived differences are discussed, and their performance is evaluated using ground truth obtained from the subjective experiment. In particular, a solution based on calibrated digital single-lens reflex camera and subsequent analysis of image features captured from the projection screen is described. The system based on our previous work is further developed so that it can be used for the analysis of projected images. It allows assessing color differences in these images and predict their impact on the perceived difference in image look.

Keywords: Digital cinema, film digitization, image restoration, image quality assessment, color grading, perceived color differences.

1. INTRODUCTION

The film archives around the world maintain since the end of the nineteenth century their film collections resulting in vast amounts of moving images captured using traditional photochemical process and now waiting to be digitized. Since the first film scanners became available, the film archives started with the digitization of their collections. However, the equipment for scanning, processing, and storage of digitized movie capable of capturing film images without severe loss of information appeared only recently. The methodologies used in the digitization were at first rudimentary, and the obtained results were far from perfect. Sophisticated techniques for digitization, especially for the movies with high artistic value, are being developed since the beginning of the twenty-first century together with the digital cinema. Nowadays, movie theaters equipped with film projectors are relatively rare, most of the current cinemas are built around digital architecture specified by the Digital Cinema Initiatives (DCI).[1] After solving the technical issues, film professionals and enthusiasts started to search for the methodology to achieve optimal results in the process of digitization. One of the methodologies being developed is the Digitally Restored Authorizate (DRA).[2,3] In the following two subsections the DRA concept is at first briefly introduced and then techniques for the subjective and objective assessment of perceived color differences among various outcomes of color grading process are discussed, forming an introduction to the actual technical contribution of this paper.

Further author information: (Send correspondence to Karel Fliegel)
Karel Fliegel: E-mail: fliegek@fel.cvut.cz, Telephone: +420 224 352 026
Stanislav Vítek: E-mail: viteks@fel.cvut.cz, Telephone: +420 224 352 232
Petr Páta: E-mail: pata@fel.cvut.cz, Telephone: +420 224 352 248
Petr Janout: E-mail: janoupe3@fel.cvut.cz, Telephone: +420 224 352 113
Jiří Myslík: E-mail: jiri.myslik@famu.cz, Telephone: +420 234 244 326
Josef Pecák: E-mail: josef.pecak@famu.cz, Telephone: +420 234 244 326
Marek Jícha: E-mail: marek.jicha@famu.cz, Telephone: +420 234 244 323

Applications of Digital Image Processing XXXIX, edited by Andrew G. Tescher, Proc. of SPIE
Vol. 9971, 997121 · © 2016 SPIE · CCC code: 0277-786X/16/$18 · doi: 10.1117/12.2237558

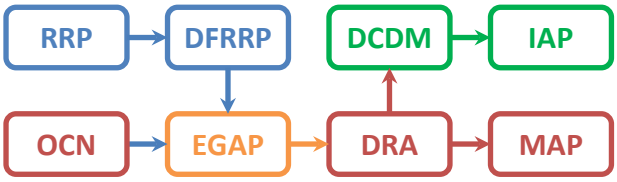


Figure 1: Simplified diagram of DRA methodology.
Original sources are in red: Original Camera Negative (OCN), Digitally Restored Authorizate (DRA), Master Archive Package (MAP). Copies are in blue: Reference Release Print (RRP), Digital Facsimile of Reference Release Print (DFRRP). Digital dissemination masters are in green: Digital Cinema Distribution Master (DCDM), Intermediate Access Package (IAP). Orange color denotes the crucial step of restoration based on estimation of the DRA utilizing the Educated Guess of Answer Print (EGAP).[2,3]

1.1 Methodology for digital restoration of archive films

The restoration methodology resulting in the DRA aims to achieve the appearance of the digitized film so that the audio and the visual component is as close as possible to the author's original concept as presented at its premiere. To achieve this goal, the digitization of the visual part of the movie should be performed by an expert restorer and supervised by a group of experts, which is composed following the specified rules and its members are professional cinematographers. Their effort should lead to the preservation of important image characteristics of the original film, i.e. color and light tonality. The goal can be achieved if the digitization follows the DRA methodology, especially the steps briefly reviewed in this section. A complete description of DRA methodology is beyond the scope of this paper.[2,3] The simplified diagram describing the main idea can be seen in Fig. 1.

At first, a group of experts is selected by the appointed master restorer. The group includes experts, e.g. film historian, film technologist, experienced conventional photochemical film restorers, but mainly professional cinematographers, and authors of the movie if possible. The film print with the appearance closest to the original, here denoted as Reference Release Print (RRP), is selected among the surviving prints in the film archive. The most suitable, unfortunately not always available, is the Answer Print (AP) print signed by the author for the presentation at the premiere, so called signed release print. Then reference sample scenes are selected with specified characteristics critical for the film appearance, e.g. darkest and brightest scene, average scene, daytime, non-daytime, resulting in six scenes with an option to select three more with the total duration of about 12 minutes. The chosen samples are scanned from the RRP at least in 4K resolution (considering 35 mm film stock) and subsequently color graded. The aim of the color grading in this stage is to create an accurate digital facsimile, here denoted as Digital Facsimile of Reference Release Print (DFRRP) so that naïve observers cannot easily distinguish the difference between the RRP and DFRRP. The DFRRP acts as a starting point for the restorer who can compare it during the restoration process with the image restored from the Original Camera Negative (OCN). The next crucial step is scanning of the OCN with utmost care in proper resolution of at least 4K resulting in the Digital Source Master (DSM). The most important step is the color grading of the scanned OCN lead by the restorer and supported by the group of experts resulting in the appearance as close as possible to the original authors' film concept at the time of the premiere. It mainly means to remove all unwanted color and light tonality drifts caused by the aging of the archive film. The resulting product is called Educated Guess of Answer Print (EGAP), and it is produced from the scanned OCN using digital color grading techniques as a result of the search for the original appearance of the image corresponding to the concept of the film as intended by the authors. The EGAP then represents a reference for the final step, which is a creation of the actual DRA.

The DRA of the whole movie can be created by the digital colorist alone using the EGAP samples followed by the fine tuning and approval from the restorer and the expert group. As such the described procedure saves a considerable amount of time and funds. The resulting DRA is not a version of the original work (i.e. outcome of remastering) but its original digital source. The DRA contains removable artifacts, e.g. cue marks, dirt, hair as it must include these elements to be considered as an original source of the given film. It is important to mention that all the above described steps are performed in the postproduction halls with calibrated DCI projectors. The last stage of the methodology comprises the creation of the Digital Cinema Distribution Master (DCDM) with cleaned out artifacts to be used further for the creation of all following distribution masters

for various purposes. The Distribution Access Package (DAP) can be created in different output formats, e.g. digital cinema, television, home video, online streaming. The DRA should be stored and archived preferably using mathematically lossless JPEG2000 in a form of the Master Archive Package (MAP).[4] For more details on the methodology, please refer to the respective publications.[2,3] *

1.2 Subjective and objective assessment of perceived color differences

The general goal of the restoration can be simplified in the requirement for the creation of a digital copy which will provide the audience with the experience close to the one obtained during a hypothetical projection using the signed show print in the quality as it was at the time of its premiere. It considers digital projection while fulfilling the requirements of the DCI[1] specification and related regulations, and SMPTE recommendations.[5] Following the methodology described in Section 1.1 there is a need for techniques, which would provide quantification of perceived difference among various outcomes of the color grading process.

The quantification of perceived difference might be based on the evaluation of the subjective experiment with human observers or objective measurement done on provided digital image files or using images captured directly from the projection screen. The literature addresses the issue of perceived color differences in complex image stimuli from various points of view. Recently presented study evaluated the performance of various color difference (CD) measures for the contexts where reference and distorted images are compared in full-reference (FR) manner.[6] The authors found out that the CIEDE2000[7] and its spatial extension perform well in images distorted by black level shift and color quantization but for the whole variety of color distortion in images the tested CD measures reached rather poor correlation (PLCC - Pearson Linear Correlation Coefficient - denoted as r, SROCC - Spearman Rank Order Correlation Coefficient - denoted as ρ)[8] with subjective results $\mid \rho \mid < 0.65$. Despite this fact, most of the available techniques for evaluation of perceived color distortions in complex stimuli are based on S-CIELAB,[9] i.e. spatial extension of CIELAB color space, CIEDE2000 and its spatial extensions.[10,11] Application of CIEDE2000 for objective measurement of perceived color differences utilizing images captured directly from the screen by the digital camera was discussed in details in our previous paper, with good correlation $\mid \rho \mid > 0.90$ for a dataset with 180 images distorted using 6 different color distortion types.[12]

In this paper, we aim at further development and performance analysis of our previously presented system for objective assessment of image differences in digital cinema.[12] The previous system was developed and tested for a particular set of color distortions. Here we present results when the system is used in a more realistic scenario. Broad types of color distortions are allowed as a product of color grading process but with more subtle perceived differences than in the previously used dataset.

Evaluation of perceived color differences among the three complex stimuli was performed within the subjective experiment in the scenario where three outcomes of color grading process applied to the same film samples were available. The results of this experiment were statistically processed and further employed as a ground truth for performance evaluation of the system based on objective assessment of color differences in various configurations. The paper is divided into four main sections. Conducted subjective study is introduced, and obtained results are presented in Section 2. Description of various configurations of the system for objective assessment of color differences using a digital camera is presented in Section 3. Experimental results are presented, and the performance of the system is evaluated in Section 4. In Section 5, the conclusions are provided, and possible future work is discussed.

2. SUBJECTIVE EXPERIMENT

The goal of the subjective experiment conducted within this study was to obtain ground truth data for the perceived differences between the images obtained from independent color grading procedures while projected onto the cinema screen. A description of the conducted subjective experiment is presented in the following section. This section is divided into five subsections, where the preparation of the test content is at first discussed, followed by a description of the subjective test methodology and the laboratory environment, with the statistical analysis of obtained results presented in the last subsection.

*Digitally Restored Authorizate [retrieved August 8, 2016]: `http://www.research-dra.com/`

Table 1: Description of the test content. For each image there are three independent color graded versions available.

Scene content (basic description)	**Test images** (see Fig. 2)	**Number of images**
Sunny and average scene	S1/F1 (01), S1/F2 (02), S1/F3 (03)	3
Early evening	S2/F1 (04), S2/F2 (05), S2/F3 (06), S2/F3 (07)	4
High contrast	S3/F1 (08), S3/F2 (09)	2
Nigh with moon reflection in water	S4/F1 (10), S4/F2 (11), S4/F3 (12)	3
Candle lit scene	S5/F1 (13), S5/F2 (14), S5/F3 (15)	3
Sunny day with skin tones	S6/F1 (16), S6/F2 (17)	2
High contrast shoot against window	S7/F1 (18)	1
Dark scene with skin tones	S8/F2 (19)	1
Darkest scene with details and skin tones	S9/F1 (20)	1

2.1 Preparation of the test content

As described in the Section 1.1, proposed DRA methodology for digitization of archive films has several important steps. One of the steps is the creation of the Educated Guess of Answer Print (EGAP) for selected sample images from the scenes crucial to the final film look. The content for the subjective experiment was obtained using the procedure described in the Section 1.1 from the actual color movie "Capricious Summer" or "Rozmarné léto" in Czech, important piece of Czechoslovak cinematography produced in 1967.* It is important that for this particular movie there are the Reference Release Print (RRP) and Original Camera Negative (OCN) available.

According to the methodology, key scenes were selected, and independent color grading was performed by the three independent digital colorists, restorers, and their expert groups. The expert group selected nine common key scenes with an average duration of about 90 s and followed by the selection of 20 particular still frames from these scenes, see Tab. 1. Thumbnails of the test frames (outcome from one of the color gradings) can be seen in Fig. 2. For each of the 20 test contents, three test conditions are resulting in 60 test images. The goal of the subjective experiment is to assess the perceived difference between these three modifications for all the 20 frames. The test images were kept in their original formats to be properly handled by the color grading system for the correct projection with DCI certified digital postproduction projector. The images were also exported into 16 bpc TIFF files in RGB and XYZ color spaces for further analysis. The resolution of these images was 4096×3112 pixels (4K Full Aperture "Open Gate"), with the active image area of about 3680×2670, i.e. with the aspect ratio of about 1:1.37.

2.2 Test conditions

In our previous work[12] the subjective experiment was conducted in the laboratory environment with carefully prepared stimuli, covering a broad range of manipulations applied globally on the whole image. The distorted images were obtained from the reference ones by (a) Brightness, (b) Contrast, (c) Color saturation, (d) Color balance modification.

The test images used in the experiment and described in this paper were not obtained by a systematical alternation of selected image features but are a result of the actual application of the restoration process described in the Section 1.1. As mentioned in the Section 2.1, there are three modifications available for each of the 20 frames obtained through the color grading process conducted by the three independent experts group. Due to this fact, it can be expected that the three modifications are very close in their appearance and the perceived color differences. Moreover, in contrast to the previous experiment, the expected color differences are composed of various manipulation types and do not cover densely the broad distortion space. An example of the Digital Facsimile of Reference Release Print (DFRRP) for the selected test content and the three modifications obtained

*Capricious Summer [retrieved August 8, 2016]: `http://www.imdb.com/title/tt0063527/`

(a) S1/F1 (01) (b) S1/F2 (02) (c) S1/F3 (03) (d) S2/F1 (04) (e) S2/F2 (05)

(f) S2/F3 (06) (g) S2/F4 (07) (h) S3/F1 (08) (i) S3/F2 (09) (j) S4/F1 (10)

(k) S4/F2 (11) (l) S4/F3 (12) (m) S5/F1 (13) (n) S5/F2 (14) (o) S5/F3 (15)

(p) S6/F1 (16) (q) S6/F2 (17) (r) S7/F1 (18) (s) S8/F2 (19) (t) S9/F1 (20)

Figure 2: Selected test contents from the movie "Capricious Summer".
Only the thumbnails of the actual test content are depicted here, with levels modified for a proper appearance in printed form. Notation S[m]/F[n] (k) stands for Scene [m]/Frame [n] (Test Content k).

as a result of the color grading process for the three expert groups can be seen in Fig. 3. It is clear that the perceived color differences between the outcome of the three groups are rather small.

2.3 Test methodology

The methodology used for the subjective experiment was based on the modified Double Stimulus Impairment Scale method (DSIS) as described in ITU-R BT.500-13[13] recommendation. The two stimuli for the outcome of the two selected groups were projected onto the projection screen in a time sequence. The viewers were asked to evaluate the perceived difference between the two stimuli on the projection screen, see Fig. 4.

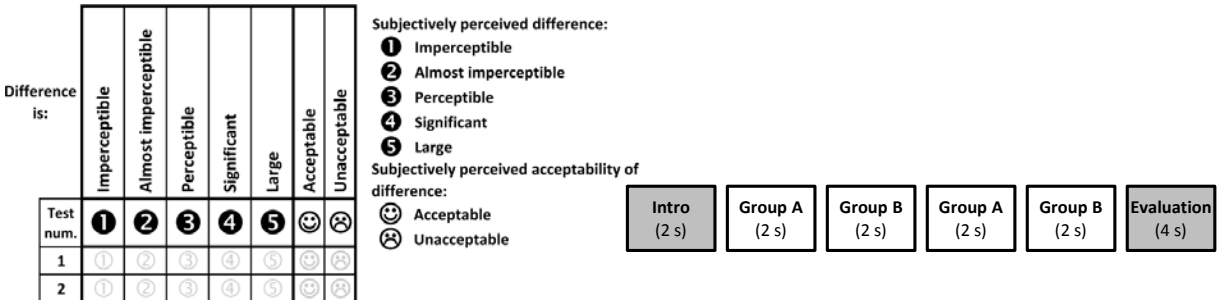
The test session is described in more details in the following paragraph. The subjects were at first familiarized with the aim of the subjective assessment, i.e. to determine the impact of the perceived difference between the stimuli. The two stimuli being compared were obtained by two independent color grading procedures, see Section 2.1 and 2.2. The observers were instructed to imagine that there is a sharp cut between the renditions of the particular scene and that their task is to assess the perceived difference between the two renditions. Then the viewers evaluated the perceived difference between the stimuli using the five-point scale, see Fig. 4(a). As an additional information, the viewers were asked to provide their opinion on the acceptability of the difference. The five point scale was proposed by the expert cinematographers. The particular wording is used in their practice to describe the level of distortion to be corrected in the post-production process of color grading. The scale and assigned wording for the particular levels of "Subjectively perceived difference" is (1) Imperceptible, (2) Almost

(a) DFRRP (w/ soundtrack) (b) Group 1 (c) Group 2 (d) Group 3

Figure 3: Comparison of the Digital Facsimile of Reference Release Print (DFRRP) of the selected test content (S6/F1 (16), Fig. 2(p)) and the three outcomes of color grading process obtained by the three independent expert groups. Perceived difference between the color appearance of the DFRRP and the outcome of the expert groups is evident, but the differences between the groups are tiny and might be unnoticeable in the printed form.

imperceptible, (3) Perceptible, (4) Significant, and (5) Large. An example of the English version of the form can be seen in Fig. 4(a).

Overall 75 pairs of stimuli were evaluated in the subjective experiment. It included all the 60 pairs (20 contents and three combinations of the stimuli for the groups 1-2, 1-3, and 2-3) and hidden 15 training pairs. The order of the pairs and the order between the stimuli was randomized. The evaluation was performed for each group of observers in one session, lasting for about 20 minutes. The presentation structure of test material, see Fig. 4(b) was set as follows: (1) mid-gray screen with a number of the test (2 seconds), (2) first stimulus obtained for the Group A (2 seconds), (3) second stimulus obtained for the Group B (2 seconds), (4) repeated steps (2) and (3), and (5) mid-gray screen with a number of the test to be judged (4 seconds). The group of observers in the subjective experiment was limited to 17 persons with the seating optimal for the viewing conditions in the projection hall. Observers were equipped with a writing board and tiny LED reading light.



(a) Evaluation scale and part of the form. (b) Phases of presentation.

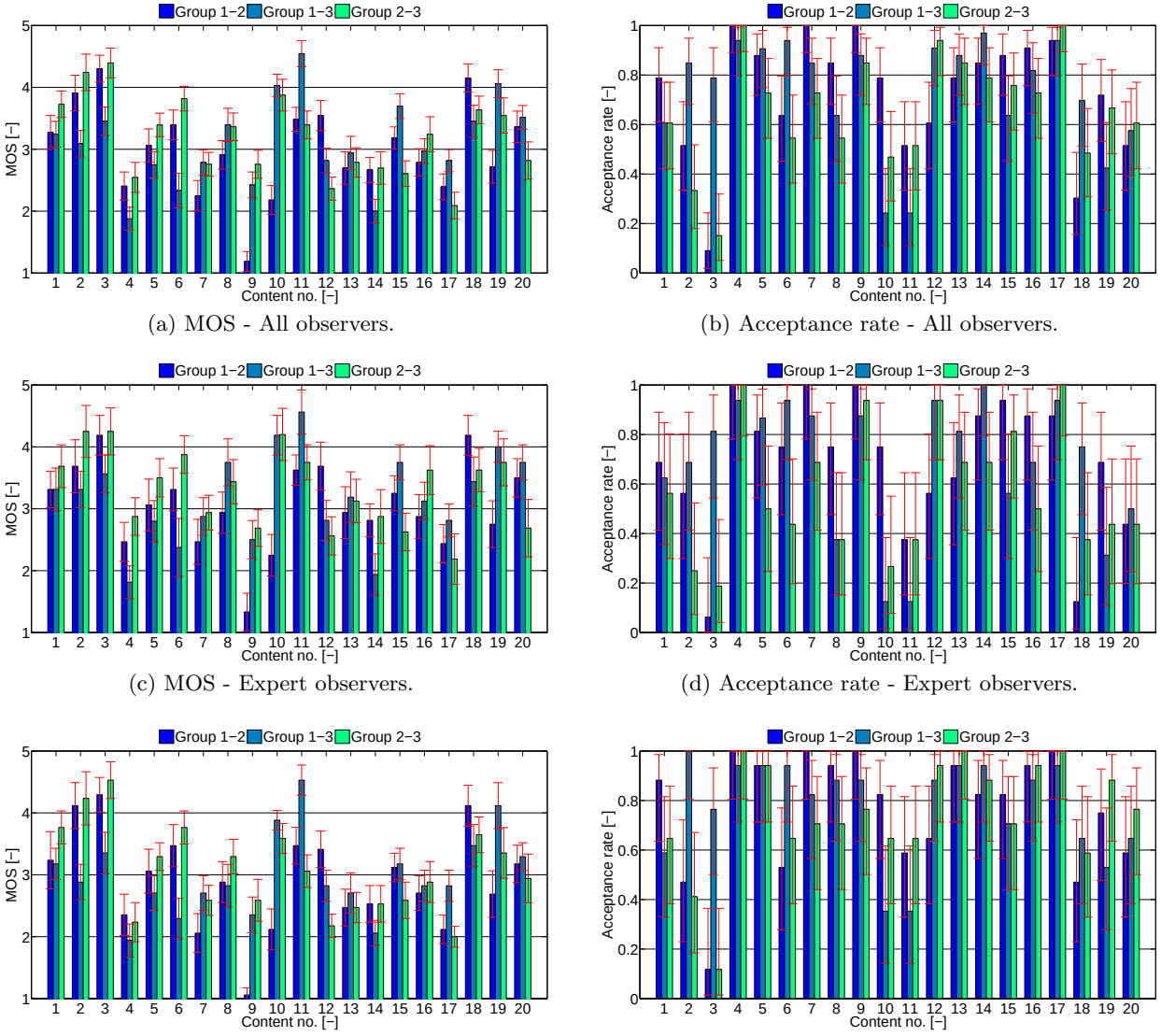
Figure 4: Subjective experiment. The evaluation form and the phases of presentation.

2.4 Presentation environment

The subjective experiment took place in the professional review hall of the Film and TV School of Academy of Performing Arts in Prague. The hall and the used equipment fulfill the requirements of SMPTE for the review rooms.[5] The room is also equipped with acoustic isolation and dampening to prevent unwanted noise and sound reflections.

The digital projector used in the experiment was Barco 4K digital posproduction projector DP4K-P* with the native 4K resolution of 4096×2160 pixels. The projector meets the high-performance demands for post-production, archiving, restoration and 4K color grading. The projector is compliant with SMPTE RP431-2[5] standard. As the primary characteristic, the luminance in the center of the projection screen should be $48\,\mathrm{cd/m^2}$

*Barco [retrieved August 9, 2016]: `https://www.barco.com/en/Products/Projectors`



(a) MOS - All observers.

(b) Acceptance rate - All observers.

(c) MOS - Expert observers.

(d) Acceptance rate - Expert observers.

(e) MOS - Naïve observers.

(f) Acceptance rate - Naïve observers.

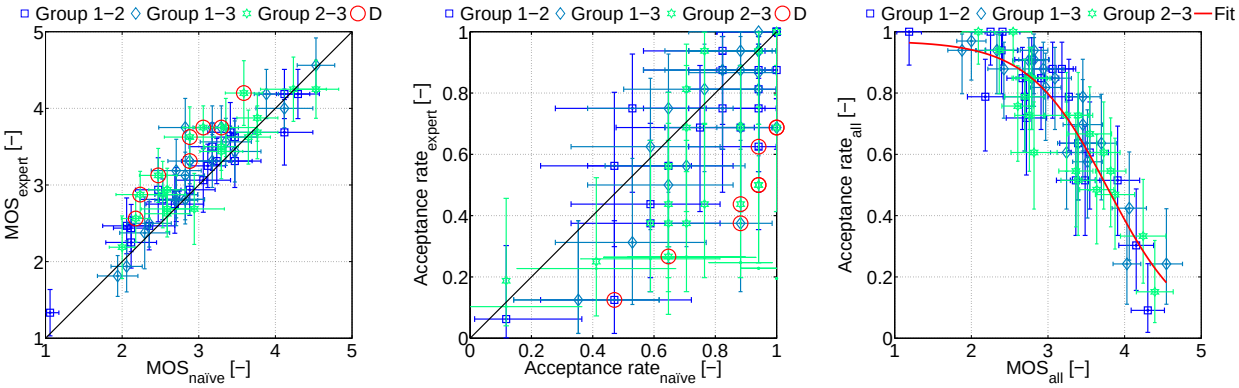
Figure 5: Results of the subjective experiment for all the contents (see Fig. 2) and test conditions, i.e. perceived difference of the color grading outcomes between the groups (Group 1-2, Group 1-3, and Group 2-3). There are MOS values in the left column and Acceptance rate values in the right column. In the first row, there are results obtained for all observers, in the second row for expert observers and in the third row for naïve observers. 95% confidence intervals are also depicted in red color.

and chromaticity $x = 0.314$, $y = 0.351$ for the white point. The projector was thoroughly calibrated, and the parameters were verified using a professional grade spectroradiometer Photo Research SpectraScan PR-740*. The obtained parameters were well within the SMPTE tolerances for the review rooms (± 3.5 cd/m^2, ± 0.002 xy).

2.5 Results of subjective experiment

The subjective experiment took place in June 2016 with 33 observers. Among the observers, there were 17 naïve (14 male and 3 female, i.e. 17.6% female), most of whom were university students and faculty, supplemented by some externals with an age range in between 22 and 78 years, an average age of 34.5 years and median age of

*Photo Research: `http://www.photoresearch.com/current/pr740.asp`



(a) MOS - expert vs. MOS - naïve (b) AR - expert vs. AR - naïve. (c) AR - all vs. MOS - all.

Figure 6: Dependencies between the MOS and Acceptance rate calculated from the scores of expert, naïve or all observers. Confidence intervals in both directions are also depicted. Statistically significant difference in the results calculated from the scores on expert and naïve observers is denoted using a red circle. The relationship between the Acceptance rate (AR) and MOS calculated for all observers is depicted in subfigure (c) with logistic fit included (red curve). Results obtained for the particular test condition, comparing the outcome of color grading between expert groups (Group 1-2, Group 1-3, and Group 2-3), are color coded with distinct markers.

25 years. The other half of the observers was formed of 16 experts (all males), most of whom were professional cinematographers with an age range in between 23 and 76 years, the average age of 50.7 years and the median age of 53 years. The observers were divided into two groups with a maximum number of 17 observers for a session. Before the experiment took place, the observers filled a general questionnaire. The ophthalmologist regularly checks the expert observers with no need for further screening, the naïve observers were not screened but did not report bad condition of their vision.

The performed statistical processing of subjective ratings recorded in paper forms was done according to the standard procedures described in ITU-R BT.500-13[13] recommendation. The screening of the observers based on their ratings was not performed because this procedure should be restricted to the cases in which there are relatively few observers, i.e. fewer than 20 and the whole panel was composed of 33 observers. Mean opinion scores (MOS) and Acceptance rate with 95% confidence intervals (CI) for each tested content and condition were calculated. The Acceptance rate quantifies the expected ratio of observers assessing the particular content and test condition (perceived color difference) as acceptable.

The results are depicted in Fig. 5 in a form of a group bar graph with assigned errorbars in red color defining the confidence intervals. There are six graphs, two for MOS and Acceptance rate calculated for the scores of all observers, expert observers and naïve observers. The confidence intervals CI for MOS are narrow, fitting easily in one step on the selected rating scale. It is good to note that the CIs calculated from the scores of expert observers are in most of the cases broader than for the naïve observers. This might be explained by the fact that the previous experience might bias the scores of the expert observers. The CIs for the Acceptance rate are much broader, which is due to the binary nature of the scoring and nature of the used Clopper-Pearson exact intervals in comparison to the classical CI calculation for normally distributed subjective scores. Detailed description of the statistical processing of the data is beyond the scope of this paper; more details can be found in the literature.[8,14]

Dependencies between the results calculated for a group of experts and naïve observers can be also analyzed in a form of scatter plot, see Fig. 6. From the relationship between MOS for expert and naïve observers it can be seen that experts tend to score the perceived color difference as more apparent. Majority of the MOS values lies above the identity line, see Fig. 6(a). However, from the point of statistical significance, the difference in MOS values is statistically significant only for eight pairs out of 60 in total. Therefore, for further analysis in this paper, MOS values calculated for the combined group of all observers were used. Obviously, in the case of

Acceptance rate (AR), the expert observers were more critical than the naïve ones. In this case, most of the AR values is under the identity line. Because of broad confidence intervals, the difference between the AR values is statistically significant only for seven pairs out of 60 in total. The significance level of 0.05 for the two-tailed test was considered. Thus, also for the AR, values calculated for the combined group of all observers were considered for further analysis.

Table 2: Table of MOS and Acceptance rate (AR) values calculated from the logistic fit based on the results of the subjective experiment (see Fig. 6(c)).

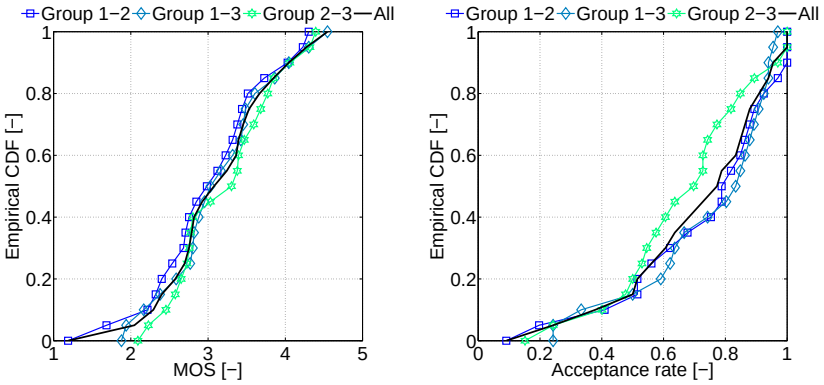
Measure	**Values obtained from logistic fit**									
MOS [-]	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5
AR [-]	0.97	0.97	0.96	0.94	0.89	0.80	0.62	0.39	0.19	0.08
AR [-]	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
MOS [-]	4.38	4.26	4.12	3.94	3.72	3.47	3.19	2.89	2.57	2.24

One can expect that there is a relationship between the MOS and Acceptance rate (AR) obtained from the subjective experiment. In other words, the Acceptance rate could be predicted from the MOS if there is some simple relationship. The prediction accuracy can be evaluated through the Pearson linear correlation coefficient (PLCC), sometimes denoted as Pearson's r. Whereas prediction monotonicity using Spearman rank order correlation coefficient (SROCC), sometimes referred to as Spearman's ρ. More details on the performance testing techniques can be found in various references.[8,15,16] Obtained relationship between the AR and MOS can be seen as a scatter plot in Fig. 6(c), together with calculated confidence intervals in both directions and fitted using logistic sigmoid curve. It is evident that there is a relationship between the AR and MOS, resulting in correlation coefficients and their 95% confidence intervals, $r = -0.88^{-0.80}_{-0.92}$ and $\rho = -0.87^{-0.76}_{-0.93}$. Obtained correlation coefficients with values approaching 0.90 indicate good prediction accuracy and monotonicity.

Based on the logistic sigmoid curve (red curve in Fig. 6(c)), relationship for selected values of AR and MOS is summarized in Tab. 2. The logistic fit shows, for example, an estimation that for MOS lower than 3 "Perceptible" is the Acceptance rate of about $AR = 0.80$, i.e. the perceived difference will be acceptable for more than 80% of observers. On the other hand for example if the perceived difference should be acceptable for more than 50% observers, then the MOS should be lower than about 3.7, i.e between the scores 4 "Significant" and 3 "Perceptible" on the scale, see Fig. 4. It is important to note that this conversion using the logistic fit is very approximate considering the content dependency (see broad CIs in Fig. 6(c)) for complex stimuli.

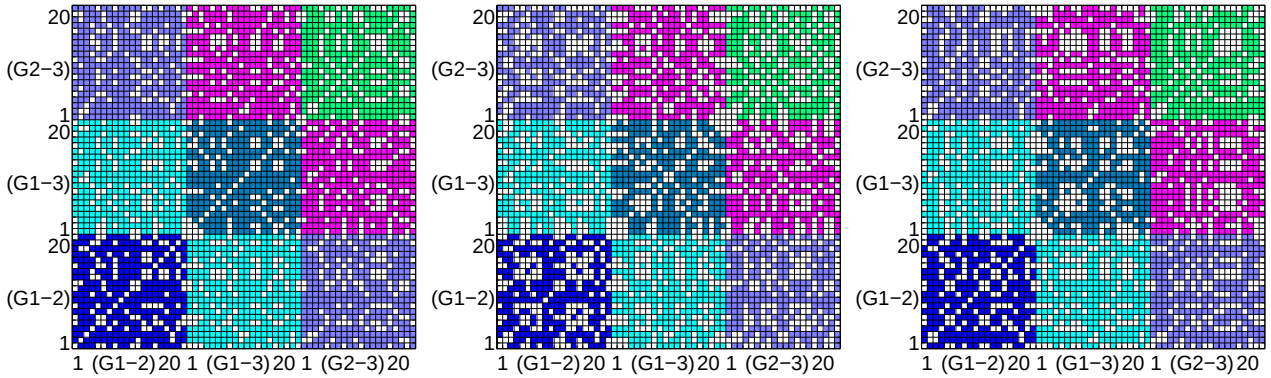
Analysis of MOS and Acceptance rate (AR) values in the particular tests can be seen in a form of empirical Cumulative Distribution Function in Fig. 7. The empirical CDF shows how the values of MOS and AR are distributed among the test pairs used in the experiment. This graphical representation of results also allows for qualitative comparison among the test conditions, here comparing the outcome of color grading between expert groups (Group 1-2, Group 1-3, and Group 2-3). It can be seen that perceived color differences in the measure of MOS values for the test condition Group 2-3 are slightly higher than for the other two test conditions, Group 1-2 and Group 1-3, see Fig. 7(a). The range of MOS and AR values obtained in the subjective experiment would cover the whole rating scale, with higher density in small differences, even if the number of test pairs was rather limited. The empirical CDF for AR can also be used to predict the percentage of the test pairs with the required or better AR. For example, if the required AR is at least 0.5, i.e. the perceived color difference is acceptable for at least 50% of observers, then the corresponding empirical CDF is about 0.15, i.e. more than 85% of test pairs would fulfill the AR requirement. If the AR requirement would be to have the perceived color difference acceptable for at least 80% of observers, then it would be fulfilled for about 42% of test pairs. These values are valid if the whole set of test pairs is taken into account. Results for the test condition Group 2-3 are worse. Here it is important to note that these estimates are only approximate due to broad CIs in the assessment of MOS and especially AR, see Fig. 6.

To verify if the differences in MOS values between the particular test pairs and test conditions are statistically significant, paired-sample two-tailed t-tests at significance level 0.05 were performed.[8,17] The results and



(a) Empirical CDF for MOS. (b) Empirical CDF for AR.

Figure 7: Empirical Cumulative Distribution Function (CDF) of tests for MOS and Acceptance rate (AR). Results obtained for the particular test condition, comparing the outcome of color grading between expert groups (Group 1-2, Group 1-3, and Group 2-3), are color coded with distinct markers. Empirical CDF for the combination of the three test conditions is plotted using black curve.



(a) All observers. (b) Expert observers. (c) Naïve observers.

Figure 8: Results of the paired-sample two-tailed t-tests for all the test pairs. The white square between two test pairs means that the difference between the two MOS ratings is not statistically significant. Particular test conditions compare the outcome of color grading between expert groups (Group 1-2, Group 1-3, and Group 2-3), and they are color coded.

calculated with ratings from all observers, expert, and naïve observers, see Fig. 8. If the square between the two test conditions and contents combinations is white, then there is no statistically significant difference between the two MOS ratings. Otherwise, the square is color coded. It is clear that there is a good ratio of test pairs with statistical significance between the respective MOS values. Based on the presented analysis, the obtained results can be used for development and testing of an objective technique for evaluation of color differences in cinematographic images similarly as in our previous paper.[12] This problem is discussed in the following section.

3. OBJECTIVE EVALUATION OF COLOR DIFFERENCES

In the previous Section 2 an analysis of the outcome of the subjective study was presented. In this section, there are selected techniques introduced with the aim reliably and objectively assess perceived color differences in projected cinematographic images. There are three different approaches proposed, and their performance is evaluated using the data from the subjective experiment as a ground truth. One of the techniques is based on capturing images directly from the projection screen using a digital camera and their subsequent evaluation. This technique is based on our previous work[12,18] and is further developed here. The second method uses a

professional spectroradiometer to measure and evaluate selected color samples directly from the projection screen. The third approach uses source image files in digital form and their colorimetric analysis.

3.1 General description of the proposed techniques

The perceived difference in images projected onto the cinematographic screen can be assessed through a subjective experiment with a group of observers, see Section 2, and objectively based on a computational comparison of the two projected images with different color or light tonality. General description of the proposed techniques is presented in this section. There were three conceptually different approaches used to obtain the colorimetric data about the images projected onto the projection screen used then to assess color differences between images objectively, based on:

(a) measurement of selected color samples directly from the projection screen using spectroradiometer,
(b) capturing the projected images using calibrated digital camera,
(c) direct evaluation of source image files.

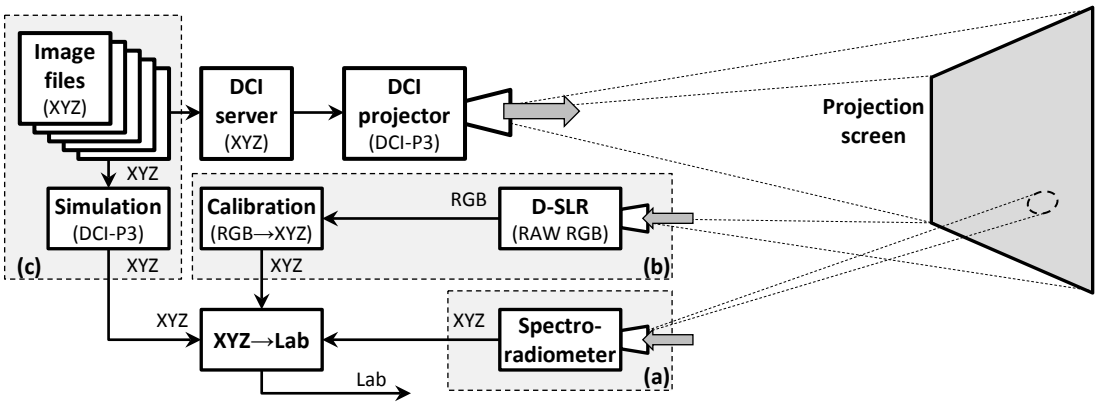


Figure 9: Block diagram describing the general principle of the three discussed techniques for derivation of colorimetric data in projected cinematographic images. System (a) is based on the measurement of CIE XYZ tristimulus values in selected color samples using spectroradiometer. System (b) captures the 2D distribution of CIE XYZ tristimulus values using calibrated digital single lens reflex camera (D-SLR). System (c) is based on the calculation of CIE XYZ tristimulus values using a simulation of digital projection.

For all the three mentioned approaches the primary goal is to obtain reliable colorimetric rendition of the picture on the projection screen in device independent color space, preferably as CIE XYZ tristimulus values.[19] The general principle of the three discussed techniques can be consulted in Fig. 9. The first approach is straightforward as the selected color samples can be directly measured from the projection screen using colorimeter or spectroradiometer, in our case professional grade spectroradiometer Photo Research SpectraScan PR-740. Thus, the measurement of the tristimulus values is very precise. The main disadvantage of this approach is that the specific key color regions have to be identified by the expert and then measured at the same spatial position for all the images being compared.

The second approach uses calibrated digital single lens reflex camera (D-SLR) to capture the test images from the screen. After the calibration, the D-SLR acts as an imaging colorimeter and provides for the captured scene a two dimensional map of CIE XYZ tristimulus values. The advantage of this approach is in the fact that there is no necessity for identification of key color samples in the evaluated images. On the other hand, the main disadvantage of this approach is the limited accuracy of measurement in comparison to the first approach. It can be solved by using a professional grade imaging colorimeter, e.g. Photo Research Tru-Image TRU8* or count with the lower accuracy of the D-SLR based solution. More details on the calibration of the D-SLR for purposes of 2D imaging colorimetric measurement can be found in the Section 3.3 and in our previous papers.[12,18]

*Photo Research [retrieved August 10, 2016]: `http://www.photoresearch.com/content/tru-image-2d-imager`

(a) S1/F1 (01), 3 samples. (b) S3/F1 (08), 4 samples. (c) S6/F1 (16), 5 samples. (d) CIE xy chroma diagram.

Figure 10: An example of the three selected contents with markers (white circles) for color sample measurement using spectroradiometer. Subfigure (d) shows CIE xy chromaticity diagram with all the 228 measured color samples depicted (black markers) and DCI-P3 color space shown (in the white triangle).

The third technique is based on the evaluation of source image data. This approach is quite straightforward because in the case of DCI image files there is a device independent representation in CIE XYZ color space directly available.[5] If there is a goal to assess differences in images projected by a particular DCI projector onto a cinematographic screen, then the image data in CIE XYZ should be at first transformed using the known colorimetric characterization of the projector. However, usually the exact characterization is not available and if yes, it often does not take into account the black level shift caused by the ambient light in the projection hall. Therefore this approach has an advantage of being the simplest without the need to capture any information from the actual projection screen if the source image data is available. The calculation of CIE XYZ tristimulus values on the projection screen is based here on simulation. However, if the aim is to assess the color differences exactly as they would be perceived in a particular projection hall, then this approach is not accurate if the colorimetric characterization of a projection system is not available. In the following paragraphs, there is a more detailed description available for each approach.

3.2 Measurement of color samples from the projection screen

The configuration of direct measurement of CIE XYZ tristimulus values from the projection screen can be seen in Fig. 9(a). Professional spectroradiometer Photo Research SpectraScan PR-740* was used to measure CIE XYZ tristimulus values for each of the 20 contents, see Fig. 2, for the outcome of the color grading for three experts groups, resulting in 60 stimuli in total. For each of the 20 contents, the experts identified key color samples within the image, ranging from two to five color samples per content.

The CIE XYZ tristimulus values were measured in the calibrated review projection hall, similar to the one described in Section 2.4, resulting in 228 measurements in total. Three selected contents with shown positions of measurement of samples can be seen in Fig. 10(a, b, c). The white markers were shown before the measurement for easy location of the spatial coordinates with a spectroradiometer. Chromaticity coordinates of measured color samples can be visualized in CIE xy chromaticity diagram, see Fig. 10(d). It is clear that the selected color samples are well within the central part of DCI-P3[1,5] color space (white triangle). Measured CIE XYZ tristimulus values were further processed to evaluate the color differences between the two projected images, see Section 3.5.

3.3 Acquisition of images from the projection screen using calibrated digital camera

Acquisition of images directly from the projection screen using calibrated digital camera was performed at the same time as the measurement using spectroradiometer, see the previous Section 3.2. Image acquisition procedure was very similar as described in our previous paper.[12] A brief overview of the camera calibration can be found in the following paragraph.

*Photo Research: `http://www.photoresearch.com/current/pr740.asp`

Camera calibration

This section only contains braief summary on camera calibration, for more details please refer to our previous papers.[12,18] Digital camera can be turned into imaging colorimeter via calibration process. It is an affordable alternative to a professional grade imaging colorimeter, an ultimate measuring device for such a purpose. Canon EOS 6D semi-professional digital single-lens reflex (D-SLR) camera was used in the experiment. It was equipped with Canon EF 24–105 mm f/4.0 L IS USM lens*.

The system with a D-SLR camera should be capable of measuring the CIE XYZ[19] tristimulus values with high accuracy and with high spatial resolution in the entire field of view. There are various approaches published on camera characterization[18,20] and analyzing the usability of commercial digital cameras for scientific purposes.[21] The captured scene is always projected by a projector in this particular application. Thus the calibration procedure was performed using a large set of color samples projected on the screen. The color samples depicted were measured on the screen using spectroradiometer Photo Research SpectraScan PR-740, providing spectral radiance, photometry and chromaticity characteristics of each sample. Then test images were captured and corresponding 16 bpp TIFF images were obtained from 14 bpp RAW image files using dcraw utility.† A transformation matrix between the camera's color space is computed as a linear least-squares regression problem.[21] For the overall performance evaluation the CIEDE2000[22] average color difference ΔE_{00} can be calculated‡. The obtained average CIEDE2000 for Canon 6D camera is $\Delta E_{00} = 0.99$. The value is low, close to the JND. Thus the calibration can be considered as very accurate. The D-SLR is thus capable of capturing CIE XYZ tristimulus values precisely and performing objective measurements of the images captured from the projection screen.

Preprocessing of evaluated images

All the images were projected on the screen from the DCI server, see Fig. 9. Set of 60 images (20 contents, 3 test conditions) was captured by Canon 6D digital camera at the exposure time of 2 s, which was selected to cover optimally the camera's dynamic range. The RAW RGB values were extracted using dcraw utility and a simple bilinear color demosaicing technique[23] was used to obtain RGB image array in the device dependent color space. Then this array was corrected by a flat-field image obtained using standard procedures.[24] The camera output was then converted into a color-corrected device-independent image in CIE XYZ color space[12,18] for further processing, see Section 3.5.

3.4 Direct processing of source image files

The third approach does not rely on measurement of images from the projection screen but works only with the source image files, see Fig. 9(c). The test images, see Fig. 2, were projected by the DCI projector controlled from the DCI server during the subjective experiment, see Section 2. The image files were exported to 16 bpc TIFF files with components in CIE XYZ color space. Numerical values in CIE XYZ color space recorded in the image file are not the same as the real physical measures as obtained using spectroradiometer or calibrated digital camera, see Sections 3.2, and 3.3.

The whole DCI color management is very well documented and thus it is possible to estimate the physical CIE XYZ tristimulus values as they would appear on the screen if the source image files would be projected using DCI certified review projector. A description of the necessary transformation procedures can be found in the literature.[1,5] In the simplified method utilized in this paper, the code values available from the image file, usually denoted as X'Y'Z', were transformed into output referred CIE XYZ tristimulus values, considering gamma conversion function with $\gamma = 2.6$. Since the exact color characterization matrix of the used projector was not available, the output referred CIE XYZ tristimulus values were used directly as an estimate of the real physical CIE XYZ values on the screen. The color samples were well within the DCI-P3 color space, see Fig. 10(c), thus the estimate is quite accurate. On the other hand, it is important to note, that this approach does not compensate the screen black level, which might be a source of moderate error in the estimation. The obtained estimation of CIE XYZ was used in further processing, see Section 3.5.

*Canon 6D at DPReview [retrieved August 9, 2016]: `https://www.dpreview.com/reviews/canon-eos-6d`

†David Coffin's dcraw [retrieved August 9, 2016]: `http://www.cybercom.net/~dcoffin/dcraw/`

‡CIEDE2000 Color-Difference [retrieved August 9, 2016]: `http://www.ece.rochester.edu/~gsharma/ciede2000/`

3.5 Selected color difference objective measures

In the previous paragraphs, three techniques were described for the measurement of CIE XYZ tristimulus color values in images projected onto the cinema screen. The primary goal of this paper is to find a method for objective comparison of perceived color differences between two images of the same content but with different color appearance on the screen. The most common method for comparison of two homogeneous color samples in perceptually uniform manner is CIEDE2000 color difference formula.[22] For comparison of complex color stimuli, i.e. color images, spatial extensions of various color difference formulas were proposed. Most of the available methods are full-reference (FR) image quality (distortion) assessment techniques (IQA) where the reference and distorted images are available.[8,15,16]

One of the techniques is based on the spatial extension of CIE Lab color space and subsequent difference calculation in this space, called S-CIELAB.[9] In S-CIELAB, spatial filtration is performed using Contrast Sensitivity Function (CSF) of Human Visual System (HSV) adapted separately for each channel. The more recent approach uses the S-CIELAB representation as an input to CIEDE2000 color difference formula providing better perceptual uniformity.[7] The technique based on CIEDE2000 color difference formula was successfully adopted for IQA of color images.[10] Based on the preliminary results presented in our paper[12] and some more recent studies on the performance of color difference measures,[6] spatial extension of CIEDE2000 color difference formula was selected as a good performing measure to assess perceived color difference between two projected images. In our approach, see Fig. 9, the images or color samples in CIE XYZ device independent color space are transformed to CIE Lab. Then the CIEDE2000 color difference formula is applied to obtain color difference map (spatial distribution of ΔE_{00} values), which is averaged to get the overall $\Delta\bar{E}_{00}$ measure. Spatial filtration, to mimic CSF of the HVS, is simplified in our approach to uniform gaussian low pass filtering. Spatial weighting can be used to take into account the higher importance of key color regions. The spatial weighting was based on the coordinates of key color samples selected by the experts, see Fig. 10 and it was applied on the color distortion maps obtained from the source image data (see Fig. 9(c)) and from the images captured by the calibrated digital camera from the screen (see Fig. 9(b)).

In the case that the CIE XYZ values are captured from the screen using calibrated digital camera it is necessary to deal with possible geometric distortion between the two captured images. In our experiment, for the sake of simplicity, the camera was perfectly fixed on a tripod to prevent such geometric distortions and thus image registration was not necessary.

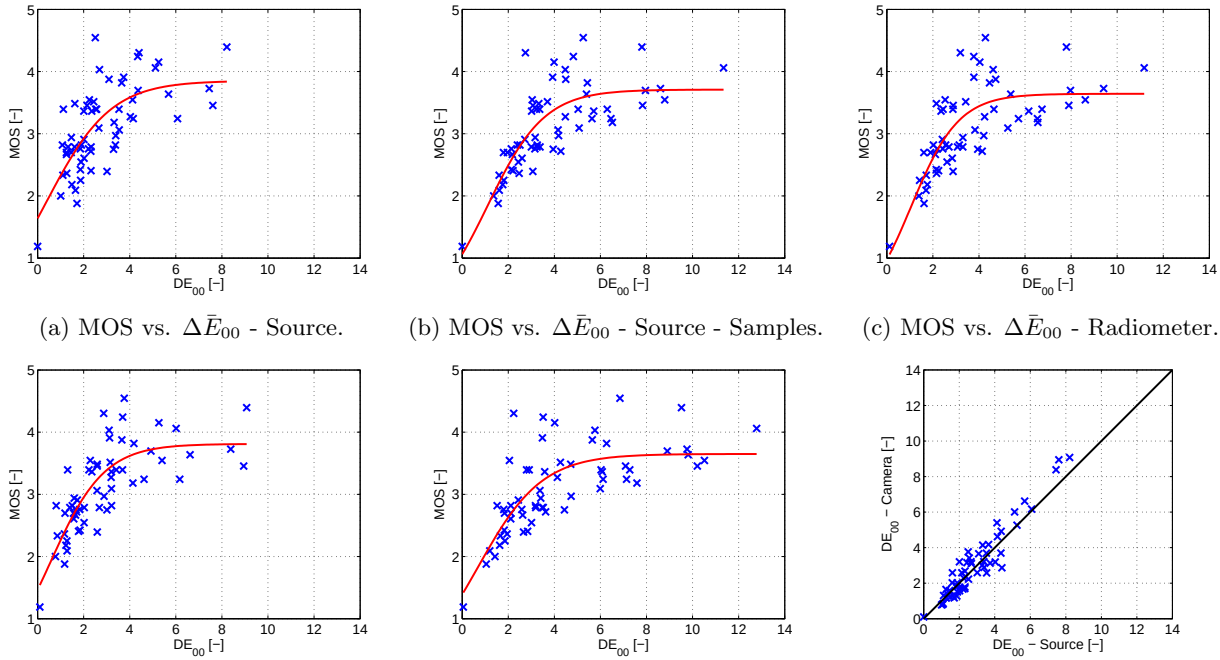
4. EXPERIMENTAL RESULTS

In this section, a performance analysis of the objective assessment techniques for prediction of perceived color differences measured through CIEDE2000 color difference formula is presented. MOS values obtained from the experiment with human observers, see Section 2, are compared with the output of objective techniques described in the Section 3.

As it was described in the Section 2.1, for each of the 20 contents (see Fig. 2) there are three test conditions (see Fig. 3) resulting in 60 image pairs. For each pair CIEDE2000 color difference formula, see Section 3.5, is used to obtain the overall measure $\Delta\bar{E}_{00}$ of perceived color difference. Each of the 60 image pairs has also corresponding MOS value assigned, see Fig. 5. Comparison of the MOS values and $\Delta\bar{E}_{00}$ values is depicted in the scatter plots, see Fig. 11(a-f). Moreover, there is a nonlinear fit based on logistic function included in the scatter plots.[8] There is also a comparison of $\Delta\bar{E}_{00}$ values obtained from the calibrated camera and the source image files.

Interesting and useful is an analysis of the color difference map of ΔE_{00} values obtained from the application of CIEDE2000 color difference formula on the data obtained form the source image files or captured from the screen by the calibrated digital camera. An example of the color difference map for one particular content and the three test conditions can be seen in Fig. 12. The map is convenient for quick assessment of the areas with larger color differences and can act as a supporting tool for a color grader.

There are various metrics used to evaluate the performance of an objective measure. Prediction accuracy can be evaluated using the Pearson linear correlation coefficient (PLCC) denoted as r, whereas prediction monotonicity through the Spearman rank order correlation coefficient (SROCC), referred to as Spearman's ρ.[8,15,16]



(a) MOS vs. $\Delta\bar{E}_{00}$ - Source. (b) MOS vs. $\Delta\bar{E}_{00}$ - Source - Samples. (c) MOS vs. $\Delta\bar{E}_{00}$ - Radiometer.

(d) MOS vs. $\Delta\bar{E}_{00}$ - Camera. (e) MOS vs. $\Delta\bar{E}_{00}$ - Camera - Samples. (f) $\Delta\bar{E}_{00}$ - Camera vs. $\Delta\bar{E}_{00}$ - Source.

Figure 11: Scatter plots of MOS versus spatially averaged CIEDE2000 color difference measure $\Delta\bar{E}_{00}$, with logistic fit included. $\Delta\bar{E}_{00}$ values were calculated using input data from spectroradiometer "Radiometer", from source image files "Source", and from the calibrated digital camera "Camera". (f) Scatter plot of $\Delta\bar{E}_{00}$ values was obtained from the source image files and the calibrated camera.

Performance metrics calculated for the tested objective measures can be found in Tab. 3. From the obtained results it can be seen that the prediction accuracy in measures of PLCC is not very good, in the range of $r = [0.619, 0.672]$. It could be partially improved if a nonlinear conversion of the measured values would be performed. On the other hand, the prediction monotonicity values using SROCC are better, in the range of $\rho = [0.702, 0.781]$. Obtained r and ρ values are also depicted in a bargraph, see Fig. 13. The prediction accuracy and monotonicity is achieved for the color differences calculated from the images captured by the calibrated digital camera from the screen.

Table 3: Performance of CIEDE2000 color difference measure $\Delta\bar{E}_{00}$ by means of PLCC r and SROCC ρ.

Measure	**Radiometer** Color samples	**Source files** Entire image	**Source files** Color samples	**Camera** Entire image	**Camera** Color samples
PLCC r	0.619	0.639	0.672	0.668	0.650
SROCC ρ	0.722	0.702	0.760	0.781	0.730

5. CONCLUSIONS AND FUTURE WORK

This paper is focused on the important topic of archive films restoration and proposes methodologies for subjective and objective assessment of perceived color differences in images on the projection screen. Subjective experiment with a group of observers allowed evaluation of the perceived differences in the outcome of the color grading process performed by three independent expert groups. Statistical analysis shows that these three outcomes are very close and the perceived color differences are not significant. The results of the subjective study were used as

(a) Group 1-2, Data. $\Delta\bar{E}_{00} = 2.32$

(b) Group 1-3, Data. $\Delta\bar{E}_{00} = 3.39$

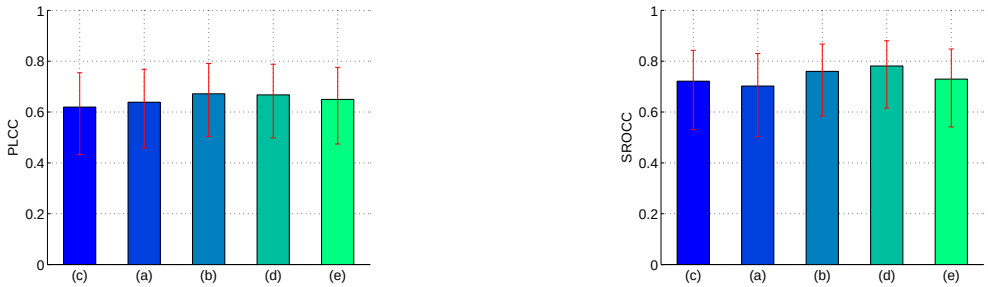
(c) Group 2-3, Data. $\Delta\bar{E}_{00} = 4.15$

(d) Group 1-2, Screen. $\Delta\bar{E}_{00} = 2.69$

(e) Group 1-3, Screen. $\Delta\bar{E}_{00} = 2.89$

(f) Group 2-3, Screen. $\Delta\bar{E}_{00} = 4.62$

Figure 12: Example of calculated maps of color difference ΔE_{00} values (displayed in the range $\Delta E_{00} = 0 - 10$). The maps are depicted for the selected test content and the three test conditions. The three outcomes are compared considering the color grading process obtained by the three independent expert groups (S6/F1 (16), see Fig. 3(b, c, d)), denoted as "Group 1-2", "Group 1-3", and "Group 2-3" in the caption. The maps on the first row were calculated directly from available image files data (denoted as "Data" in the caption) of tested image pairs and the second row was obtained using images captured by the calibrated digital camera from the projection screen (denoted as "Screen" in the caption). Spatially averaged color difference measure $\Delta\bar{E}_{00}$ is also provided.



(a) Pearson Linear Correlation Coefficient r.

(b) Spearman Rank Order Correlation Coefficient ρ.

Figure 13: Pearson Linear Correlation Coefficient (PLCC) r and Spearman Rank Order Correlation Coefficient (SROCC) ρ calculated for the MOS versus spatially averaged CIEDE2000 color difference $\Delta\bar{E}_{00}$ with confidence intervals included. Each bar is denoted using the same letter (a)-(e) as the scatter plots in Fig. 11(a-e).

a ground truth for performance analysis of the three discussed objective techniques for computational estimation of perceived color differences.

Three techniques for objective assessment of perceived color differences were developed and tested. The three methods provide a colorimetric characterization of the test images in device independent CIE XYZ color space. Two of methods are based on the analysis of the images from the projection screen using spectroradiometer or calibrated digital camera. The third method directly processes the source image files. Color differences

were quantified using a spatial extension of CIEDE2000 color difference formula. The performance of the three methods by measures of prediction accuracy and monotonicity is not ideal and a maximum of the Spearman Rank order Correlation Coefficient (SROCC) is reached at $\rho = 0.781$ for evaluation of image data captured by the calibrated camera (D-SLR) from the screen. This result shows that even if the calibrated D-SLR has lower accuracy than the spectroradiometer, this issue is compensated by better spatial characterization of the test image than by using only few color samples. Similarly, the performance of the D-SLR based approach was slightly better than the direct utilization of source image files. It is because in the latter case the simulated colorimetric characterization of the projection process brings inaccuracies.

It can be concluded that the proposed techniques for objective assessment of perceived color differences in projected images can be used to provide basic information on the visibility or acceptability of color differences. On the other hand, if an ultimate answer on the perceived color difference is required, then the subjective experiment provides more accurate data. There is a lot of space left for improvement in this area. In our future research, we will focus on further development of CIEDE2000 based techniques, more suitable for assessment of complex image stimuli.

ACKNOWLEDGMENTS

This work was supported by the project NAKI DF13P01OVV006 "Methodics of digitizing of the national film fund" of the Ministry of Culture of the Czech Republic. The authors would like to thank Lukáš Krasula for consultations on the design of the subjective experiment and all the participants involved in the subjective experiment for their time and cooperation.

REFERENCES

[1] "DCI Specification, Version 1.2 with Errata." Digital Cinema Initiatives (2012).

[2] Jícha, M. and Šofr, J., "Digitální restaurování památek filmového umění. Metoda DRA," *Zprávy památkové péče* **76**(1), 76–90 (2016). (in Czech).

[3] Jícha, M. et al., "Methodology of Digital Film Restoration Producing Digital Restored Authorizate (DRA)." Czech Society for Quality (2016). (in Czech).

[4] "ISO/IEC 15444-1:2004/Amd 2:2009 Extended profiles for cinema and video production and archival applications." International Organization for Standardization (2009).

[5] "SMPTE Standard RP 431-2:2011 D-Cinema Quality - Reference Projector and Environment." The Society of Motion Picture and Television Engineers (2011).

[6] Ortiz-Jaramillo, B., Kumcu, A., and Philips, W., "Evaluating color difference measures in images," *2016 8th International Conference on Quality of Multimedia Experience, QoMEX 2016* (2016).

[7] Johnson, G. and Fairchild, M., "A Top Down Description of S-CIELAB and CIEDE2000," *Color Research and Application* **28**(6), 425–435 (2003).

[8] Wu, H. and Rao, K., [*Digital Video Image Quality and Perceptual Coding*], Signal Processing and Communications, Taylor & Francis (2005).

[9] Zhang, X. and Wandell, B., "A spatial extension of CIELAB for digital color-image reproduction," *Journal of the Society for Information Display* **5**(1), 61–63 (1997).

[10] Yang, Y., Ming, J., and Yu, N., "Color image quality assessment based on CIEDE2000," *Advances in Multimedia* **2012** (2012).

[11] Baxter, D., Cao, F., Eliasson, H., and Phillips, J., "Development of the I3A CPIQ spatial metrics," *Proc. SPIE* **8293**, 829302–829302–12 (2012).

[12] Fliegel, K., Krasula, L., Páta, P., Myslík, J., Pecák, J., and Jícha, M., "System for objective assessment of image differences in digital cinema," *Proc. SPIE* **9217**, 92170I–92170I–14 (2014).

[13] "Rec. ITU-R BT.500-13 Methodology for the subjective assessment of the quality of television pictures." International Telecommunication Union (2012).

[14] Brown, L., Cai, T., and DasGupta, A., "Interval estimation for a binomial proportion," *Statistical Science* **16**(2), 101–133 (2001).

[15] Winkler, S., [*Digital Video Quality: Vision Models and Metrics*], Wiley (2013).

[16] Wang, Z. and Bovik, A., [*Modern Image Quality Assessment*], Synthesis lectures on image, Morgan & Claypool Publishers (2006).

[17] De Simone, F., Goldmann, L., Baroncini, V., and Ebrahimi, T., "Subjective evaluation of JPEG XR image compression," *Proc. SPIE* **7443**, 74430L–74430L–12 (2009).

[18] Fliegel, K. and Havlín, J., "Imaging photometer with a non-professional digital camera," *Proc. SPIE* **7443**, 74431Q–74431Q–8 (2009).

[19] Sharma, G. and Bala, R., [*Digital Color Imaging Handbook*], Electrical Engineering & Applied Signal Processing Series, Taylor & Francis (2002).

[20] Barnard, K. and Funt, B., "Camera characterization for color research," *Color Research & Application* **27**(3), 152–163 (2002).

[21] Akkaynak, D., Treibitz, T., Xiao, B., Guerkan, U. A., Allen, J. J., Demirci, U., and Hanlon, R. T., "Use of commercial off-the-shelf digital cameras for scientific data acquisition and scene-specific color calibration," *Journal of the Optical Society of America A: Optics and Image Science, and Vision* **31**(2), 312–321 (2014).

[22] Sharma, G., Wu, W., and Daa, E., "The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations," *Color Research & Application* **30**(1), 21–30 (2005).

[23] Lukac, R., [*Single-Sensor Imaging: Methods and Applications for Digital Cameras*], Image Processing Series, Taylor & Francis (2008).

[24] Buil, C., [*CCD Astronomy: Construction and Use of an Astronomical CCD Camera*], Cognitive Development, Willmann-Bell (1991).