

The complex process of image color correction: a test of a target-based framework

GABRIELE SIMONE,^{1,*} MARCO GAIANI,² ANDREA BALLABENI,² AND ALESSANDRO RIZZI¹

¹*MIPS Lab - Computer Science Department - University of Milan, Italy.*

²*Dipartimento di Architettura - Università di Bologna, Italy.*

gabriele.simone@di.unimi.it

Abstract: This paper aims at presenting the complexity of process of image target-based color correction (CC). We present issues encountered from acquisition to rendering using colorimetric traditional tools. Target-based CC can be seen as an optimization problem. We have tested SHAFT (SAT & HUE Adaptive Fine Tuning) an automated framework for target-based CC. A key element of SHAFT is an iterative CIEDE2000 variation comparison between a reference and target image. In this work we replace the standard CIEDE2000 with the Euclidean color-difference formula for small–medium color differences in log-compressed OSA-UCS (Optical Society of America's committee on Uniform Color Scales) space. Results are presented using both formulae. A discussion on the complexity of scene color departures and correction performances concludes the paper. It is shown the effect of real scene complexity and how colors are subject to disordered shifts in the color space. Because of this complexity, it emerges the role of the CC method as a different color error minimizer.

© 2020 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

1. Introduction

Following a theoretical approach, one could think that correcting color in an image can be seen as finding a kind of unique shift in a well-ordered series of color difference among patches. Real implementations are very far from this ideal situation. Many different factors form a long list of potential departures from the value of the scene to acquire, to the point that the many patches of a target inserted in the scene for color correction have a sort of chaotic range of color shift vectors. It is well known that differences in vector direction make the color shift more visible and annoying [1].

Usually, color device targeting is considered as the problem to solve, but real-world acquisitions are subject to many non-linearities and local variations. Among the many factors that affect the acquisition of a scene, lens glare is the one usually responsible of the biggest departure [2]. Even if rarely measured, the difference introduced by glare can easily exceed 100 % in the dark areas [1,2,3,4]. Its characteristic is to affect areas differently, following also their mutual spatial distribution. This add a series of local changes that cannot be accounted by a global color Look-Up Table (LUT). In fact, different points in the scene with the same radiance values can easily end up in different acquired values due to different distances from brighter areas in the scene arrangement. In any case, color correction remains a useful step to perform, and target-based color correction is a diffused method to compensate unavoidable color differences between the scene and its digital image. This happens in many fields of digital photography like e.g., portraiture, fashion, furniture design, interior design; among them some aims at documenting color like e.g., Architectural Heritage and Cultural Heritage.

In this paper we want to present and test a color correction method, despite the unavoidable limits of the acquisition.

All the departures from the theoretical value of acquisition, makes the target-based color correction an interpolation issue, where the practical impossibility of linear coherent shifts

gives importance to the minimization of vectorial color differences. Many open issues exist, besides the general problem of the mathematical shape of the transformation to map device-dependent and device-independent spaces with adequate performance (linear transformations, multidimensional LUTs, least-squares polynomial regressions, and others), making results sometimes unreliable. Moreover, many works consider color differences as a scalar value, while it is known how the vectorial direction of the color difference is crucial to link it to its actual perceived visual difference (ΔV) [1], especially with the need to interpolate multiple color patches, with color shift likely in different directions.

In this paper we present an example of real-world acquisitions showing the above-described problems. We used a color correction technique called SHAFT (SAT & HUE Adaptive Fine Tuning) [5], an automated framework for target-based ‘Color Correction’ (CC) benchmarked successfully in many and different use-cases mainly related to the Cultural Heritage as that in [6]. In SHAFT the optimization process is guided in the original formulation by an iterative CIEDE2000 variation comparison. In this work is presented a variant of the framework replacing the CIEDE2000 with the Euclidean color-difference formula for small–medium color differences in log-compressed OSA-UCS (Optical Society of America's committee on Uniform Color Scales) space [7], [8]. Our attempt has been inspired by the study [9], where the UCS diagram based on CIELUV color space is used to evaluate two color difference formulae ΔE_{00} and ΔE_E for measuring the visual data. Authors report that statistically, there is no significant difference between the Euclidean ΔE_E and the non-Euclidean CIEDE2000 formulae compared to the BFD-P data set [10], however, the performance of the CIEDE2000 is found weaker in blue and red region compared to the ΔE_E .

2. A short review of the SHAFT technique

The SHAFT is a software for target-based CC supported by RawTherapee, a cross-platform open-source RAW image processing program (<https://github.com/Beep6581/RawTherapee>) to which are entrusted the operations of demosaicing, devignetting, white balance and gamma correction. SHAFT is based on the ‘Color Correction’ (CC) linear approach by Bruce Fraser, the so-called Adobe Camera Raw (ACR) calibration scripts for calibration by iterative approximations. As target are admitted the X-Rite ColorChecker Classic (XCC) (Fig. 1), Passport, and Digital SG ColorChecker. The software differs from the original technique for the number and the types of tests done along the processing pipeline and for the algorithm used to find the best variation from the original values of the selected parameters (exposure, contrast, white balance, hue and saturation on each RGB channel). SHAFT is completely written in MATLAB to overcome the problem of ACR scripts that are built on Adobe Photoshop. The software is completely automated, exploiting a previously proposed solution for the target recognition on the image.



Fig 1. The target X-Rite ColorChecker Classic.

To avoid its main limitation (i.e., original highly incorrect images with high chromaticity) SHAFT is coupled with a polynomial regression correction based on least squares fitting, a solution widely adopted by many color researchers for calculation of the transfer matrix from the captured *RGB* values and their reference *XYZ* values [11]. Used references spectral reflectance values come from Danny Pascale paper [12] for the XCC used before 2016, and from measurements accomplished using a Minolta CM-2600d spectrophotometer for the XCC used after 2016. *L*a*b** to *XYZ* transformations were accomplished using MathWorks MATLAB function *lab2xyz* where the *wp* variable was set at D65 illuminant. Polynomial regression is simple and effective and in our case is achieved using a per-channel *Polynomial* curve fitting algorithm, the MATLAB *Weighted Polyfit (x,y,n)* function that returns the coefficients for a polynomial $p(x)$ of degree n . Goodness of fit coefficient and polynomial degree are identified using a weighted fit minimizing the RMSE:

$$RMSE = \sqrt{\sum_{i=1}^n \omega_i (\hat{x}_i - x_i)^2} \quad (1)$$

where ω is the weight of each patch of the XCC. The degree of the polynomial, to increase the robustness of the results, has been fixed to two. In our pipeline only the basic processing was retained: bad pixel removal; dark frame, bias subtraction, and flat-field correction; green channel equilibrium correction; Bayer interpolation. To avoid uncontrolled modification of the RAW pixel intensity values we did not allow these on-camera processes: black point subtraction; denoising; color scaling; image sharpening; color space conversion; gamma correction; format conversion. Our automatic workflow is described in [13] and is as follows:

1. RAW image 16-bit linearization and devignetting
2. Image denoise
3. Color target detection
4. Exposure equalization and white balance
5. Polynomial function for CC
6. Image CC using the new fitting function
7. SHAFT CC
8. Image color rendition using the selected color space

In detail, white balance was performed on the patch 22 of the XCC (Fig. 1), performing a simple von Kries-type transformation in *XYZ* color space. As a white point the D65 illuminant was selected. The post camera flat fielding process was performed according to [14] before the image denoising.

3. The color difference formula in target-based CC

In evaluating color image capture, it is normally a validation effort aimed at determining goodness of the color-correction operation. This involves comparing the target colorimetry to that predicted from the color profile-processed pixel values. Several visual color difference formulae are used to do this. Their use is not limited to the evaluation of the deviation from the desired capture of color image information but is part of a calibration function to establish the best possible mapping from camera *RGBs* values to device independent *XYZs* values (or rendered *RGB* values). A common measurement technique for small color differences is the CIEDE2000 color metric, termed ΔE_{00} computed for each color patch in the CIELAB color space given by the CIE in 1976 [15]:

$$\Delta E_{00} = \sqrt{\left(\frac{\Delta L'}{k_L S_L}\right)^2 + \left(\frac{\Delta C'}{k_C S_C}\right)^2 + \left(\frac{\Delta H'}{k_H S_H}\right)^2 + R_\tau \frac{\Delta C'}{k_C S_C} \frac{\Delta H'}{k_H S_H}} \quad (2)$$

The CIEDE2000 formula, recommended by CIE mainly for color differences within the range 0-5 CIELAB units [16], is the improved version of the CIE Euclidean distance measure in the

CIELAB color space. In the formula $\Delta L'$, $\Delta C'$, and $\Delta H'$ are the CIELAB metric lightness, chroma, and hue differences, respectively, calculated between the standard and sample in a pair. Specific weighting functions known as lightness (S_L), chroma (S_C) and hue (S_H), parametric factors (k_L, k_C, k_H), and the rotation term R_t have been added to correct to chroma and hue differences in the blue region. The values calculated for these functions vary according to the positions of the sample pair being considered in CIELAB color space. The relative importance of these five corrections to CIELAB is not the same: it is found that the weighting function for chroma is the most important one from a quantitative point of view [17], while the proposed weighting function for lightness seems to be relatively controversial [18]. Recent papers show improvement to the CIEDE2000 formula (e.g., by simple power functions as in [19]; or setting $S_L = 1$ as in [20]), achieving results better about the agreement with visually perceived color differences (ΔV) (i.e., the color difference as judged by an experienced person, or by a panel of observers with non-defective color vision, between two color samples - typically two homogeneous patches, in the simplest case-). However, the statistical differences reachable are marginal [21] and then the original formulation remains easier and more reliable. Finally, it is important to note that CIEDE2000 has not an associated color space [22] and then, potentially, the use of a more perceptually uniform color space as the OSA- UCS system could be a solution allowing major improvement in the visual results of the calculations.

4. Euclidean color-difference formula for small–medium color differences in log-compressed OSA-UCS space

The OSA-UCS is a color order system developed between the years 1947-1974 from an idea of Judd and the consecutive works of MacAdam [23]. The intention of the OSA committee was to show equally perceptible differences between all pairs of adjacent colors and proposed a set of three orthogonal coordinates (L_{OSA}, j, g) arranging the color samples according to a cube-octahedron lattice, which represent lightness, yellowness, and greenness, respectively [24]. The OSA-UCS system is defined for the CIE 1964 observer and D65 illuminant. The color difference formula developed for this color system, termed ΔE_E , is Euclidean and was proposed by Oleari et al. 35 years after the original work [7]. To achieve a Euclidean color difference formula Oleari et al. proposed a set of appropriate transformations of the original coordinates L_{OSA}, g, j to have a set of new coordinates L_E, G_E, J_E in the so-called log-compressed OSA-UCS space. We recall all of them now.

Given X_{10}, Y_{10}, Z_{10} and x_{10}, y_{10}, z_{10} , L_{OSA} is defined as follows:

$$L_{OSA} = \left[5.9(Y_0^{1/3} - \frac{2}{3} + C) - 14.4 \right] \frac{1}{\sqrt{2}}, \quad (3)$$

with:

$$C = \begin{cases} +0.042|Y_0 - 30|^{1/3} & \text{for } (Y_0 - 30) > 0 \\ -0.042|Y_0 - 30|^{1/3} & \text{for } (Y_0 - 30) \leq 0 \end{cases} \text{ with } Y_0 = Y_{10}F, \quad (4)$$

with:

$$F = 4.4934x_{10}^2 + 4.3034y_{10}^2 - 4.2760x_{10}y_{10} - 1.3744x_{10} - 2.5643y_{10} + 1.8103. \quad (5)$$

From Eqs. 3-4 we can derive L_E as follows:

$$L_E = \left(\frac{1}{b_L} \right) \ln \left(1 + 10 \left(\frac{b_L}{a_L} \right) L_{OSA} \right), \quad (6)$$

with:

$$a_L = 2.89; b_L = 0.015. \quad (7)$$

The coordinates G, J are defined as follows:

$$\begin{pmatrix} J \\ G \end{pmatrix} = \begin{pmatrix} 2(0.5735L_{OSA} + 7.0892) & 0 \\ 0 & -2(0.7640L_{OSA} + 9.2521) \end{pmatrix} \begin{pmatrix} 0.1792 & 0.9837 \\ 0.9482 & -0.3175 \end{pmatrix} \begin{pmatrix} \ln\left(\frac{A/B}{0.9366}\right) \\ \ln\left(\frac{B/C}{0.9807}\right) \end{pmatrix}$$

, (8) with:

$$\begin{pmatrix} A \\ B \\ C \end{pmatrix} = \begin{pmatrix} 0.65973 & 0.44916 & -0.10889 \\ -0.30528 & 1.21255 & 0.09273 \\ -0.03740 & 0.47951 & 0.55789 \end{pmatrix} \begin{pmatrix} X_{10} \\ Y_{10} \\ Z_{10} \end{pmatrix} \quad (9)$$

where X_{10}, Y_{10}, Z_{10} are the tristimulus values of the sample.

From Eq. 8 we can derive G_E, J_E as follows:

$$G_E = -C_E \cos(h), J_E = C_E \sin(h), \quad (10)$$

with:

$$h = \arctan\left(-\frac{J}{G}\right), C_E = \left(\frac{1}{b_c}\right) \ln\left(1 + 10 \left(\frac{b_c}{a_c}\right) C_{OSA}\right)$$

$$a_c = 1.256, b_c = 0.050, C_{OSA} = \sqrt{G^2 + J^2}. \quad (11)$$

Finally, we can define the Euclidean color-difference formula for small-medium color differences in log-compressed OSA-UCS space as follows:

$$\Delta_E = \sqrt{(\Delta L_E)^2 + (\Delta G_E)^2 + (\Delta J_E)^2} \quad (12)$$

Lower complexity than CIEDE2000 and equally good in the prediction of many available empirical datasets [23] have invited the formula to be encapsulated tested in the SHAFT framework described above.

5. Experimental setup

Image Quality Assessment of color corrected images processed using the two different color spaces (CIELAB and log-compressed OSA-UCS space) was made exploiting both objective and subjective methods. Tests results of the new solution using a common reference, the target XCC [25] are presented showing that a lower number of iterations is required to reach the same numerical performance. Experiments with users are presented to evaluate the quality of the framework. As final rendered color space we used the sRGB-IEC 61966-2-1 defined with respect to CIE illuminant D65. Downsides of the sRGB color space (gamma value approximately of 2.2; colors gamut narrower than the human one, not allowing to display properly saturated colors such as yellow cadmium and blue cobalt, the color of the patch F3 of the XCC clipped) are well balanced by the 100% displayability on today's consumer monitors and the full support from the 3D Graphic API enabling a faithful reproduction of color in 3D applications. For the experiment, our dataset of 22 XCC RAW images described in [13] was used (Fig. 2). The dataset consists of images captured in photo studio, indoor, architectural scenarios, and outdoor environment, where natural light characteristics are extremely complex and changeable and represent different cases and problems of target-based CC. These targets present many types of distortions: noise, glare, uneven illumination typically of the real scene capture, but in the dataset are also target captured in a photo studio environment using professional cameras. These distortions need to be added to the inaccuracies typical of the technique, well surveyed by [26] that demonstrated that between the many signal elements that modifies the real camera responses (e.g., ringing, blurring, glare, noise, etc.), the operator errors in the capture process, and the problem related to camera orientation with respect to the target and the lighting sources, deviations can largely be attributed to operator error.



Fig 2. Dataset of 22 RAW images evaluated.

Consistency in the capture process and uniform illumination of the target are the primary requirements. The uniformity can be mathematically compensated for by capturing an image of a diffuse white board in the location of the test target using flat-fielding techniques. Vignetting could generate problem just in the case of target covering the whole area of the image which practically never happens. Typical problems of the real scene capture are instead the color cast appearing in the case of environment with surrounding walls or objects with saturated colors and, more important, the glare appearing when the camera is not well oriented with respect to the target. Moreover, Lanaro et al. [27] noted that glare is generated by lenses producing an incorrect scaling of grey patches in captured images. This acquisition distortion affects the subsequent CC. The absolute value of the measurement of the ΔE_{00} of each target reported in Table 1 well represent the number of distortions presents in the current image.

Objective assessment was performed on the images rendered in the sRGB color space, using these image quality factors calculated using the Imatest Master software (<https://www.imatest.com>):

1. ΔE_{00} mean excluding the patch F3 which color is outside of the sRGB color space
2. ΔL mean of lightness
3. Exposure error in f-stops measured by pixel levels of patches B4-E4, using gamma values measured rather than the standard value for the color space (i.e., in the case of sRGB, 2.2)
4. Processing time
5. Number of cycles of the SHAFT algorithm.

Obviously, SHAFT improve over a traditional ‘manual’ CC. In Appendix 1 are reported results of a ‘manual’ CC against a SHAFT-based CC on three selected images to see how well that compares colorimetrically to the reference and the SHAFT processing.

The subjective image quality assessment method is based on ISO 20462, of which Part 1 is an overview of practical psychophysical components [28]. In practice, we used the psychophysical technique of paired comparisons [29], performing two experiments:

- In the Experiment A (for preference), two versions of the same image processed in the two-color spaces were presented along with the original (reference) XCC placed in a box. The XCC was positioned within a light booth sized 538x310x273 mm built for this experiment colored as the F4 patch of the XCC. The light sources used are two high-quality LED Relio² (<https://www.relio.it>) illumination devices emitting continuous spectrum light at a CCT of 6500K, a neutral white with high color rendering, an illuminance of 40000 lux at 0.25 m, and CRI(R_a)>95%. Nineteen images among the dataset of 22 XCC were presented to the observer. Three images have been discarded because the size of the target was too small for a visual judgement.

- In the Experiment B (for accuracy) the image of a single patch with its reference XCC placed in a box is presented to the observer. To avoid an excessive observer's fatigue, the number of trials has been reduced selecting randomly 4 targets among the 19 images. For each target, a subset of 9 patches has been chosen, for areas where CIE color space is particularly compressed. A total of 72 patches (36 for each color space) were presented to the observer. An evaluation of the accuracy of the reproduction of each color patch using a categorical judgment was requested.

For the Experiment A we used a 'forced-choice' comparison - 3AFC (three-alternative forced choice) - displaying both the tests images and reference for a fixed amount of time (15 seconds). After that, images disappear from the screen and observers will be asked to choose the image with higher quality. Observers are always required to choose one image even if both images possess no difference. There is no time limit for observers to make the decision. We used this approach based on three items, instead of the most common based on two elements, fundamentally because the triplet comparison experiment can be judged more quickly than the three separate paired comparisons as stated in Annex D of the ISO 20462-2 [28].

In the Experiment B the subjective assessment was done using a *categorical rating* with a mean opinion score techniques applied to a double-stimulus experiment. The observer is asked to rate the quality of the single patches on an abstract scale containing one of the five categories: excellent, good, fair, poor, or not worth keeping [30] [31], mapped to numbers between 5 (excellent) and 1 (Obvious difference, which is not acceptable). The observers were enabled to move a slider allowing a precision of the judgement of 0,01 to allow high granularity. Observers then were encouraged to report the results in decimal values, e.g., a color difference of category 3.8 could be described as a good correspondence. Finally, the *Z-score* technique is used instead a simple mean opinion score (the arithmetic mean over single ratings performed by human subjects for the given stimulus) to avoid unreliable results, e.g., caused by the fact that observers are likely to assign different quality scales to each scene and even distortion types [32].

The Experiment B has been extended, and analysis were made with greater accuracy to explore the most problematic patches, the ones which colors are in the more compressed in the CIELAB color space, and then being characterized by a smaller area in the MacAdam's ellipses, typically in the blue colors and, less, of the reds. We selected 7 patches from the XCC, typically blue or bluish, namely the patches 3, 5, 6, 8, 10, 13, 22 (see figure 1). In this case each pair of images was assessed three times by each observer. In total, 168 assessments (7 patches × 2 techniques × 4 targets × 3 replications) per observer were accumulated. Each observer was encouraged to finish one session in 20 min, but there was no time limit for visual assessment for a pair of images. A total of 2016 judgments (12 observers × 168 assessments) were obtained. These repeated results were used to report intra- and inter-observer variability. Intraobserver variability or repeatability error is a deviation among replicated observations of a given observer in an experiment, and interobserver variability or observer accuracy can be considered as an average deviation between individual results and the mean results of all observers.

For the Experiment B as the *Z-score* calculated from the observation data are normally distributed, we can utilize classical parametric statistics in the further analysis. To inquire the significance of the input images, the experiments (accuracy and preference), and the conversions (i.e., the factors) on the observation data, it is profitable to apply the multifactorial Analysis of Variance (ANOVA) test [33], allowing to analyze the differences among group means in a sample.

Factors such as display device and lighting condition affect the results of these methods. For this reason, we used a set of recommendations for standardized testing methods for subjective image quality assessment by the International Telecommunications Union in 2008 [34]. The evaluated images were displayed on a BenQ PhotoVue Photographer Professional monitor with In-Plane Switching (IPS) technology, a 27-inch LED display, in native resolution of 3840x2160 pixels. It features a resolution of 163 dpi, luminance of 350 nit, and color gamut 100% sRGB. During the experiment, the monitor was calibrated daily by an X-Rite I1 display

pro sensor driven by X-Rite software i1Profiler profiling software to D65 and a gamma of 2.2 ($\Delta E_{00} = 1.8$, average value for the sample set of the XCC). We increased the peak luminance of the sRGB color space from the suggested 80 nit to 160 nit to reflect current capabilities of LCD displays rather than an average peak brightness of a CRT. Also, the spatial uniformity of the monitor was investigated. It was found that the mean ΔE_{00} at nine considered spatial locations was 0.83, with a maximum of 1.86. This result is acceptable considering that the detection threshold for assessing image difference in previous literature is >2 CIELAB units [35].

Fig. 3 illustrates the experimental setup. The images were assessed by 12 non-expert observers who were confirmed to have normal or corrected to normal vision, following the observation that currently, researchers on color differences consider that reliable results require at least 10 individual observers, with a minimum of 3 replications per observer [36].

The observers were both male and female between the ages of 26 and 67, with normal color vision, according to the Ishihara test. Prior to the experiment, all observers were trained to assess image difference using the category judgment method. After each visual assessment, the gray background was presented for a few seconds so that the observers recovered from visual adaptation to avoid the afterimage effect before the next visual assessment.

For additional reliability, all observers repeated each experimental session three times in three days, to reduce a possible learning effect. The single experimental session took approximately 30 minutes per observer. To build the visual presentation and to record users' answers, we utilized the Python library PsychoPy (<https://www.psychopy.org/>), a framework to develop psychophysics experiments developed by University of Nottingham. The displayed size of the image was set to real size of the XCC (273 x 192 mm). The two images in a pair had a separation of about 10 mm and the background was set to the XCC patch F4. The sequence of images and the position of images on the display (left or right) were randomized. The type of the experiment (accuracy or preference) was also randomized, however for a given observer it remained constant.



Fig. 3. The experimental environment. Experiment A. On the left side, the target corrected with the two methods are presented on the screen; on the right side, there is the physical XCC placed in the box.

6. Test results

Table 1 shows the results of the objective test. For each image and method, the measurements listed at Par. 5 were performed. The confidence level to compare processing time and number of cycles with T-Test was set at 95%. The two approaches based on ΔE_{00} and ΔE_E do not show a statistically significant difference in performance. 3D plots and images acquired, color corrected and using the two techniques are in Appendix 2.

Tables 2-4 show results related to the Experiment A and Experiment B. In detail, Table 2 illustrates observers' preferences of Experiment A. Table 3 presents the mean opinion score per observer of Experiment B. Finally, Table 4 presents the mean opinion score per single patch of

Experiment B. We remark that research conducted on determining the smallest mean opinion score difference perceptible to users for digital photographs demonstrated that it is approximately 0.46 and is required the 75% of the users to be able to detect the higher quality image [37]. From these results emerges that no difference appears both in objective and subjective tests between the use of different color spaces for CC. Moreover, the perceived (ΔV) and the computed (ΔE) color differences are proportional to each other.

Table 1. Objective test results. For each image and method (ΔE_{00} -based and ΔE_E -based), the measurements listed at Par. 5 were performed, using the dataset of 22 XCC RAW images illustrated in fig. 2.

Image id	ΔE_{00}							ΔE_E						
	ΔE_{00}	ΔL	$\Delta Chroma$	ΔHue	Exposure error	Processing Time	Number of cycles	ΔE_{00}	ΔL	$\Delta Chroma$	ΔHue	Exposure error	Processing Time	Number of cycles
_48_R01_	2.93	1.36	3.58	2.68	0.17	162.00	125	3.33	1.16	4.11	3.31	0.17	158.00	125
_DSC0134	2.17	1.27	2.16	1.58	0.08	31.25	125	2.31	1.09	1.76	2.05	0.08	35.01	125
_DSC0136	2.57	1.21	2.96	2.38	0.08	32.29	110	3.41	1.12	3.19	3.73	0.08	36.07	125
_DSC0945	4.39	2.79	5.12	3.15	0.18	51.00	125	5.23	2.73	5.48	4.31	0.18	51.01	30
_DSC2118	2.32	1.53	2.05	1.95	0.04	38.86	125	2.68	1.29	1.82	2.58	0.03	40.04	125
_DSC2416	2.70	1.50	3.03	2.21	0.10	37.64	220	2.85	1.30	3.80	2.53	0.10	36.08	315
_DSC2524	1.80	0.94	2.11	1.62	0.06	36.05	125	2.04	0.80	2.20	2.10	0.06	40.14	125
_DSC2801	2.01	0.98	1.98	2.02	0.06	37.40	220	2.38	0.73	2.15	2.76	0.06	40.41	220
_DSC2935	3.60	2.36	3.78	2.43	0.10	39.44	126	3.71	2.13	4.47	2.73	0.11	39.71	125
_DSC3197	2.51	1.59	2.08	2.18	0.07	33.21	505	5.47	3.62	4.82	5.43	0.07	34.24	410
_DSC3630	2.30	1.26	2.04	2.07	0.09	35.28	505	2.98	1.54	2.81	2.74	0.09	34.76	410
_DSC6062	2.35	1.62	2.81	1.66	0.15	37.84	30	2.84	1.45	2.91	2.47	0.15	40.16	220
_MG_9935	2.60	1.64	2.97	1.82	0.08	42.01	30	3.26	1.74	3.55	2.67	0.08	44.71	125
_T4A4626	4.69	4.15	3.50	2.56	0.08	47.48	30	5.11	4.00	3.62	4.02	0.08	49.92	125
_T4A4846	2.36	1.20	3.16	2.03	0.11	48.18	125	5.11	4.00	3.62	4.02	0.08	50.48	31
_T4A5900	2.73	1.26	3.23	2.42	0.11	45.27	220	3.00	1.41	2.83	2.89	0.06	48.49	125
_wb_L5_A	2.33	1.32	2.42	1.97	0.08	48.00	200	3.00	1.10	2.36	3.18	0.08	48.00	200
1_200Is	1.98	0.96	2.23	1.86	0.08	50.00	125	2.24	0.69	2.00	2.47	0.08	52.44	125
2_200Is	2.46	1.24	3.91	2.11	0.08	50.73	31	2.84	1.02	3.95	2.69	0.08	53.39	126
3_200Is	2.20	1.16	2.29	1.72	0.09	50.11	30	2.78	0.70	2.03	3.12	0.09	50.01	125
APC_0029	3.70	2.20	5.09	2.60	0.14	35.13	126	3.86	2.17	5.54	2.61	0.14	36.32	125
APC_0033	5.42	3.39	6.34	3.97	0.19	31.45	315	5.93	3.21	7.11	4.68	0.18	37.69	220
Mean	2.82	1.68	3.15	2.24	0.10	46.39	162.41	3.47	1.77	3.51	3.19	0.10	48.05	167.36
T Test on processing time: 0.356617002							T Test on number of cycles: 0.889405395							

Table 2. Number of XCC image preferences for each observer.

Observer	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	Total
ΔE_{00}	11	10	13	8	10	6	10	12	10	7	11	12	120
ΔE_E	8	9	6	11	9	13	9	7	9	12	8	7	108

Table 3. Mean opinion score corrected with the *Z-score* technique per each observer on the 72 patches (36 for each color space). Average ΔE_{00} and ΔE_E scores are represented as *Z-scores* for each experimental subject.

Observer	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	Total
ΔE_{00}	-0.42	-0.27	-0.43	0.5	0.58	0.23	0.42	0.11	0.22	-0.66	0.56	-0.53	0.03
ΔE_E	-0.53	-0.3	-0.49	0.49	0.59	-0.21	0.26	0.12	0.03	-0.51	0.83	-0.58	-0.03

Table 4. Mean opinion score corrected with the *Z-score* technique of all observers per each patch (36 for each color space). Average ΔE_{00} and ΔE_E scores are represented as *Z-scores* for each experimental subject.

Patch name	1	7	13	14	3	15	16	11	17	Total	
ΔE_{00}		-0.26	0.21	-0.77	-0.18	-0.64	0.16	0.33	0.77	0.63	0.03
ΔE_E		-0.39	-0.2	-0.8	-0.28	-0.33	0.17	0.15	0.76	0.62	-0.03

Table 5. One-Way ANOVA. Factor is represented by the different development, based on ΔE_{00} and ΔE_E . MS represents the mean of square, DoF represents the Degrees of Freedom, F is Fisher value, F_{crit} represents the critical value of F distribution. The two groups represent the ΔE_{00} and ΔE_E methods. The significance level obtained (0.480845) is below 0.05 ($\alpha = 0.05$), meaning that there is no significant difference between the means of the two groups.

ANALYSIS OF VARIANCE						
Origin of variation	MS	DoF	MS	F	Significance level	F_{crit}
Between groups	0.448427	1	0.448427	0.497346	0.480845	3.851445
Within groups	841.229	933	0.901639			
Total	841.6774	934				

However, analyzing in-depth areas in the CIEDE diagram where colors are compressed (i.e., the blue colors area) the differences between the two-color spaces result greater and we can state that the use of the log-compressed OSA-UCS space could give better results. Shifts expressed in raw values plotted in 3D RGB-space and in white patch balanced chromaticity values, are reported in Appendix 3.

Tables 6-7 show results related to the Extended Experiment B. Table 6 presents the mean opinion score per observer of Extended Experiment B. Finally, Table 7 presents the mean opinion score per single patch of Extended Experiment B. The analysis of the blueish patches showed a statistically significant difference between color patches developed using ΔE_E and ΔE_{00} as a correction criterion. This difference has been highlighted using one-way ANOVA. ANOVA tested two hypotheses: the null hypothesis (H_0) assumes equal average scores within groups. The alternate hypothesis (H_1) assumes a different average in at least one group. To test the null hypothesis H_0 , ANOVA uses a statistical test, known as F test [38]. F test is the result of the ratio between two variances:

$$F = \frac{SSB}{k-1} / \frac{SSW}{N-k} \quad (11)$$

where SSB is the sum of square between groups, $k-1$ are the degrees of freedom of SSB, SSW the sum of square within groups, N is the number of observations, and $N-k$ are the degrees of freedom of SSW. The chosen significance level is 5% ($\alpha=0.05$), assuming a 5% error in accepting erroneously the null hypothesis H_0 . *Z-score* have been divided in two groups according to their development strategy: the first group was related to images developed with ΔE_{00} correction strategy and the second one to images developed with ΔE_E correction strategy.

The results of the ANOVA are summarized in two tables [39] (Tables 5 and 8). Table 5 does not show significant differences between the average *Z-score* of the two groups, while Table 8 shows statistically significant differences in the average *Z-score* of the two groups. In

the latter case, the significance level is 0.014405, which is less than 0.05. This result suggests that the performances of the conversions depend on input images and on experiment type, and it makes sense to show the results separately for each input image and for each experiment.

Table 6. Extended Experiment B - Mean opinion score corrected with the Z-score technique per each observer on the 168 patches observed (84 for each color space). Average ΔE_{00} and ΔE_E scores are represented as Z-scores for each experimental subject.

Observer	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	Total
ΔE_{00}	0.17	-1.24	0.52	-0.46	0.27	0.39	0.72	0.17	0.22	0.35	-0.6	-0.73	-0.02
ΔE_E	0.22	-1.28	0.52	-0.49	0.3	0.44	0.73	0.17	0.31	0.61	-0.61	-0.7	0.02

Table 7. Extended Experiment B - Mean opinion score corrected with the Z-score technique of all observers per each patch (84 for each color space). Average ΔE_{00} and ΔE_E scores are represented as Z-scores for each patch.

Patch name	13	3	17	5	6	10	8	Total	
ΔE_{00}		-0.32	-0.64	-0.2	0.03	0.09	0.87	0.05	-0.02
ΔE_E		-0.26	-0.65	-0.1	-0.01	0.09	0.95	0.1	0.02

Table 8. Extended Experiment B - One-Way ANOVA. Factor is represented by the different CC solutions, based on ΔE_{00} and ΔE_E . MS represents the mean of square, DoF represents the Degrees of Freedom, F is Fisher value, F_{crit} represents the critical value of F distribution. The significance level obtained (0.014405) is below 0.05 (alpha = 0.05).

ANALISYS OF VARIANCE						
Origin of variation	MS	DoF	MS	F	Significance level	F_{crit}
Between groups	7.241202	1	7.241202	5.997383	0.014405	3.845724
Within groups	2634.533	2182	1.207394			
Total	2641.774	2183				

7. Discussion

Replacing the traditional CIEDE2000 with a simpler, more recent comparable color difference formula, such as the Euclidean color-difference formula for small-medium color differences in log-compressed OSA-UCS, is just one step in the way to improve results. In the development and optimization of a target-based CC framework, other crucial steps need to be considered. CC in practice does not overlap with theory. This is because colorimetry does not consider many aspects that influence the results [40] e.g., illuminants are never uniform in real scene and cause interactions and interreflections among objects. Besides, also gamut and gamut mapping, hidden camera computations, RAW vs jpg non-linearities, ringing, blurring, vignetting, and many other phenomena should be considered. Along with these, also different types of noise may occur, like e.g., dark current noise.

A last but not least factor is glare, a systematic departure due to light spread through the lens. Its magnitude can be very high and cause a big departure from scene values, especially for dark regions.

A quantitative analysis is performed on the differences between CIEDE2000 and ΔEE . Such analysis has been conducted comparing a synthetic set of color chart acquisitions. All images have been captured during daytime, and thus a light source very close to D65 illuminant has been considered for CIELAB computation. A simple gaussian distributed noise and a simple general offset have been added to a synthetic raw XCC to mimic glare. Table 9 and Fig. 4 report

the results showing the influence of glare and noise on both techniques. Glare and noise affect the results in a similar way for both the two experimented techniques.

Table 9. Quantitative analysis of the influence of increasing glare and noise on the two techniques based on ΔE_{00} and ΔE_E measured using ΔE_{00} units

Glare	Noise	ΔE_{00}			Glare	Noise	ΔE_E		
		0%	5%	10%			0%	5%	10%
0		0.07	0.1	4.67	0		0.23	0.13	5.91
10		0.83	0.96	5.03	10		0.92	1.02	6.3
20		0.96	1.41	6	20		1.02	1.55	6.25

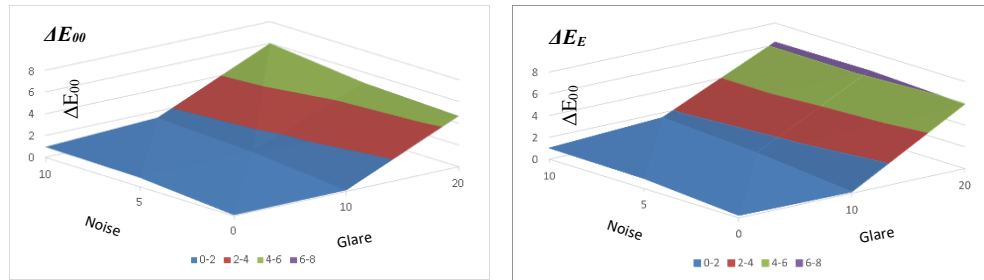


Fig. 4. Quantitative analysis of the influence of increasing glare and noise on the two techniques based on ΔE_{00} and ΔE_E . Graphical results.

Table 10. CIELAB values for patch 24. Acquired, CIEDE2000 and DEE from left to right, respectively. Reference value are: $L^*=20.64$, $a^*=0.13$, $b^*=-0.46$

Image	Acquired			ΔE_{00}			ΔE_E		
	L^*	a^*	b^*	L^*	a^*	b^*	L^*	a^*	b^*
_48_R01_	16.96	0.39	-1.34	23.06	-2.40	1.27	23.12	-2.20	2.04
_DSC0134	25.59	-1.38	-9.81	21.37	-0.60	-0.17	21.39	-0.63	-0.16
_DSC0136	6.27	0.24	1.65	19.08	-0.19	0.66	19.02	-0.25	0.43
_DSC0945	20.65	0.98	-1.89	26.69	-2.22	2.23	26.69	-2.22	2.24
_DSC2118	25.01	0.07	-5.20	22.05	-0.97	0.70	22.04	-0.97	0.62
_DSC2416	21.76	1.44	-0.69	21.53	-1.95	0.60	21.57	-1.86	0.59
_DSC2524	23.23	0.87	-2.21	20.57	0.60	0.34	20.56	0.63	0.29
_DSC2801	27.38	2.28	1.79	20.28	-0.35	-0.64	20.34	-0.41	-0.57
_DSC2935	20.67	-0.91	-4.90	25.28	0.04	1.32	25.29	0.06	1.32
_DSC3197	12.07	0.53	-5.26	20.81	-0.77	-0.43	20.90	-0.89	-0.36
_DSC3630	36.58	0.19	-6.11	21.24	-1.79	-0.29	21.35	-1.83	-0.05
_DSC6062	7.83	1.63	3.68	22.99	-0.93	0.27	22.98	-0.91	0.21
_MG_9935	18.92	0.47	-6.00	23.06	-0.83	1.59	23.03	-0.90	1.48
_T4A4626	7.82	0.27	-3.55	22.43	-1.00	1.93	22.37	-0.88	1.79
_T4A4846	16.26	2.52	-6.67	22.06	-1.12	0.60	22.37	-0.88	1.79
_T4A5900	19.53	1.31	3.97	22.05	-1.11	0.51	20.16	-2.27	0.54
_wb_L5_A	24.85	2.02	-3.03	19.35	-0.47	-0.32	19.35	-0.49	-0.40
1_200Is	23.48	-1.18	-6.70	20.06	-0.53	-0.81	20.09	-0.60	-0.77
2_200Is	25.60	-1.13	-5.18	20.04	0.21	-0.65	20.07	0.17	-0.63
3_200Is	25.65	0.50	-0.88	20.04	0.27	0.00	20.02	0.29	-0.06
APC_0029	21.53	0.38	-1.88	24.32	-1.05	1.50	24.35	-0.94	1.47
APC_0033	21.41	0.11	-0.77	26.47	-2.10	1.94	26.44	-2.29	2.49

Table 10 reports the values of the black patch of XCC in different steps of the process: as reference value, acquired directly on XCC using the Minolta CM-2600d spectrophotometer, the CIELAB value of pixels, averaged on the central part of the patch, for all images, the CIELAB value of pixels after SHAFT correction using CIEDE2000, the CIELAB value of pixels after SHAFT correction using DEE. Table 11 reports for each scene the *Lightness difference* and ΔE of the XCC black patch (#24) with respect to reference value.

Table 11. Lightness differences and ΔE for patch 24. Acquired, CIEDE2000 and DEE from left to right, respectively. In green values < 1 and in red values > 10 , respectively.

Image	Lightness differences			ΔE		
	Acquired	ΔE_{00}	ΔE_E	Acquired	ΔE_{00}	ΔE_E
_48_R01_	3.68	-2.42	-2.48	3.79	3.91	4.22
_DSC0134	-4.95	-0.73	-0.75	10.69	1.07	1.11
_DSC0136	14.37	1.56	1.62	14.52	1.95	1.89
_DSC0945	-0.01	-6.05	-6.05	1.66	7.03	7.03
_DSC2118	-4.37	-1.41	-1.40	6.45	2.13	2.08
_DSC2416	-1.12	-0.89	-0.93	1.74	2.50	2.43
_DSC2524	-2.59	0.07	0.08	3.21	0.93	0.90
_DSC2801	-6.74	0.36	0.30	7.42	0.63	0.63
_DSC2935	-0.03	-4.64	-4.65	4.56	4.97	4.98
_DSC3197	8.57	-0.17	-0.26	9.83	0.92	1.06
_DSC3630	-15.94	-0.60	-0.71	16.91	2.02	2.12
_DSC6062	12.81	-2.35	-2.34	13.55	2.68	2.65
_MG_9935	1.72	-2.42	-2.39	5.81	3.31	3.25
_T4A4626	12.82	-1.79	-1.73	13.19	3.19	3.01
_T4A4846	4.38	-1.42	-1.73	7.97	2.17	3.01
_T4A5900	1.11	-1.41	0.48	4.72	2.11	2.64
_wb_L5_A	-4.21	1.29	1.29	5.28	1.43	1.43
1_200Is	-2.84	0.58	0.55	6.98	0.95	0.97
2_200Is	-4.96	0.60	0.57	6.96	0.63	0.60
3_200Is	-5.01	0.60	0.62	5.04	0.77	0.75
APC_0029	-0.89	-3.68	-3.71	1.69	4.33	4.32
APC_0033	-0.77	-5.83	-5.80	0.83	6.69	6.94

Table 12. Lightness ratio between the white patch (#19) and the black patch (#24). Acquired, CIEDE2000 and DEE from left to right, respectively. In red values < 4 and in yellow values > 5 , respectively. The reference value is 4.61.

Image	Acquired	ΔE_{00}	ΔE_E
_48_R01_	5.61	4.12	4.11
_DSC0134	3.90	4.45	4.44
_DSC0136	7.89	4.99	5.00
_DSC0945	4.25	3.48	3.48
_DSC2118	3.99	4.30	4.30
_DSC2416	4.40	4.41	4.40
_DSC2524	4.28	4.62	4.62
_DSC2801	3.65	4.69	4.68
_DSC2935	4.59	3.73	3.73
_DSC3197	5.41	4.58	4.56
_DSC3630	2.73	4.48	4.46
_DSC6062	6.00	4.11	4.11
_MG_9935	4.62	4.11	4.11
_T4A4626	6.70	4.24	4.25
_T4A4846	5.99	4.29	4.25
_T4A5900	4.68	4.30	4.70
_wb_L5_A	3.98	4.92	4.92
1_200Is	4.20	4.75	4.74
2_200Is	3.91	4.75	4.74
3_200Is	3.90	4.75	4.75
APC_0029	4.45	3.89	3.88
APC_0033	4.02	3.52	3.52

Glare has its paramount effect on the dark areas. It follows that the black patch #24 is the place where glare effect is higher. Among the 22 scenes acquired, six images (48_R01_, DSC0136, DSC3197, DSC6062, T4A4626 and T4A4846) shows a value greater than 5 and thus a higher dynamic range. On the other side, seven images (DSC0134, DSC2118, DSC2801, DSC3630, wb_L5_A, 2_200Is, 3_200Is) show a ratio lower than 4, thus a lower dynamic range. The remaining ones are the closer to the reference 4.61. After correction (second and third column), all ratios are compressed and only four images fall below value 4 (in red).

In general, it is possible to observe noteworthy improvements for all scenes except for Image 4 and where we have a degradation in terms of lightness difference. Considering ΔE , among the 22 scenes acquired, only image labeled APC_0033 shows a value close to the reference. Five images (DSC0134, DSC0136, DSC3630, DSC6062 and T4A4626) instead show to be very far from the reference.

Table 12 reports the lightness ratio between the white patch (#19) and the black patch (#24). Among the 22 scenes acquired, six images (48_R01_, DSC0136, DSC3197, DSC6062, T4A4626 and T4A4846) shows to have a higher ratio than the reference and thus a higher dynamic range. On the contrary seven images (DSC0134, DSC2118, DSC2801, DSC3630, wb_L5_A, 2_200Is, 3_200Is) show a lower ratio. After SHAFT optimization with CIEDE2000 and DEE, 18 on 22 images show a distance less than 1 to the reference ratio.

Table 13 reports the Glare Evidence (GE) proposed by [4] and defined as follows:

$$GE = \frac{\left(\frac{L_B^* \text{acquired}}{L_W^* \text{acquired}}\right)}{\left(\frac{L_B^* \text{reference}}{L_W^* \text{reference}}\right)} \quad (12)$$

Table 13. The Glare evidence (GE). The reference values are: Black $L^* = 20.64$, White $L^* = 95.17$, B/W Ratio = 0.22.

Image	Acquired Black L^*	Acquired White L^*	Acquired Ratio B/W Ratio	GE Acquired / Reference
48_R01_	16.96	95.09	0.18	0.82
DSC0134	25.59	99.77	0.26	1.18
DSC0136	6.27	49.45	0.13	0.58
DSC0945	20.65	87.78	0.24	1.08
DSC2118	25.01	99.74	0.25	1.16
DSC2416	21.76	95.69	0.23	1.05
DSC2524	23.23	99.34	0.23	1.08
DSC2801	27.38	100.00	0.27	1.26
DSC2935	20.67	94.81	0.22	1.01
DSC3197	12.07	65.29	0.18	0.85
DSC3630	36.58	100.00	0.37	1.69
DSC6062	7.83	46.96	0.17	0.77
MG_9935	18.92	87.45	0.22	1.00
T4A4626	7.82	52.40	0.15	0.69
T4A4846	16.26	97.36	0.17	0.77
T4A5900	19.53	91.47	0.21	0.98
wb_L5_A	24.85	98.80	0.25	1.16
1_200Is	23.48	98.61	0.24	1.10
2_200Is	25.60	100.00	0.26	1.18
3_200Is	25.65	100.00	0.26	1.18
APC_0029	21.53	95.82	0.22	1.04
APC_0033	21.41	86.03	0.25	1.15

Glare is rarely considered in color correction practices, but its magnitude is usually too high to be ignored. One of the goals of this paper is underlining this effect. Glare evidence measure can be helpful in digital color management.

In conclusion, we can observe that target-based CC in real scene acquisition is subject to a complex series of color shifts and a correction method that can only aim at lowering these departures. In such complex scenario it is straightforward that the chosen color difference has

a role, but not the major one, and consequently it makes sense that the results of the two used formulae lead to similar results.

8. Conclusion

In this paper we have presented the complexity of image target-based CC when processing real-world scenes where the complexity is very high.

The main goal is to set a framework where perceptual and mathematical refinements of color distances were tested.

After introducing issues that can be encountered from acquisition to rendering using colorimetric traditional tools, we have tested SHAFT (SAT & HUE Adaptive Fine Tuning), an automated framework for target-based CC.

We have replaced the standard CIEDE2000 with the Euclidean color-difference formula for small-medium color differences in log-compressed OSA-UCS space in the iterative process of variation comparison between a reference and target image.

Computational load, measured as processing time and number of cycles needed to achieve the correction optimization, does not show a statistically significant difference in performance between the two formulae.

From the reported acquisition data, it has been shown the effect of real scene complexity and how colors are subject to disordered shifts in the color space. Because of this complexity, it emerges the role of the CC method as a different color error minimizer. It follows that the chosen color difference plays a secondary role, with similar results from the two tested formulae.

The type of acquisitions here presented have been realized without any preventative measures, so to be representative of everyday shots. Glare is demonstrated to affect every acquisition device that uses lenses, as well as all the other non-linearities due to e.g., shadows, interreflections, etc. In general, all the differences here presented and described are common in every possible acquisition.

Disclosure

The authors declare no conflicts of interest.

References

1. J.J. McCann, and A. Rizzi, *The Art and Science of HDR Imaging*, (John Wiley, 2011).
2. J.J. McCann, V. Vonikakis, and A. Rizzi, *HDR Scene Capture and Appearance*, (SPIE, 2017).
3. J.J. McCann, and A. Rizzi, "Camera and visual veiling glare in HDR images", *J. Soc. for Inf. Display* **15**(9), 721-730 (2007).
4. A. Signoroni, M. Conte, A. Plutino, and A. Rizzi, "Spatial-Spectral Evidence of Glare Influence on Hyperspectral Acquisitions", *Sensors*, **20**, 4374 (2020).
5. M. Gaiani, and A. Ballabeni, "SHAFT (SAT & HUE Adaptive Fine Tuning), a new automated solution for target-based color correction", in *Colour and Colorimetry Multidisciplinary Contributions* (2018), **XIVB**, pp. 69-80.
6. M.Gaiani and F.I. Apollonio, "Under the lens of ISLe: Leonardo da Vinci's "Landscape" drawing analysed by colourimetry", *Cultura e Scienza del Colore - Color Culture and Science Journal* **11**(2), 73-81 (2019).
7. C. Oleari, M. Melgosa, and R. Huertas, "Euclidean color-difference formula for small-medium color differences in log-compressed OSA-UCS space", *J. Opt. Soc. Am. A*, **26**(1), 121-134 (2009).
8. M. Melgosa, D. Pant, G. Simone, "Revisiting the Optical Society of America Uniform Color Scales system: past, present and future challenges", *Col. Technol.* **137**(1), 33-37 (2021).
9. D.R. Pant, and I. Farup, "CIE uniform chromaticity scale diagram for measuring performance of OSA-UCS ΔE and CIEDE00 formulas", in *3rd European Workshop on Visual Information Processing* (2011), pp. 18-23.
10. M.R. Luo, and B. Rigg, "Chromaticity-discrimination ellipses for surface colours", *Color Res. Appl.*, **11**, 25-42 (1986).
11. E.-S. Kim, S.-H. Lee, S.-W. Jang, and K.-I. Sohng, "Adaptive colorimetric characterization of camera for the variation of white balance," *IEICE Trans. on Electronics*, **E88-C**(11), 2086-2089 (2005).
12. D. Pascale, "RGB Coordinates of the Macbeth ColourChecker", The BabelColour Company, 2006.

13. M. Gaiani, F.I. Apollonio, A. Ballabeni, and F. Remondino, "Securing Color Fidelity in 3D Architectural Heritage Scenarios", *Sensors* **17**(11), 2437-1-2437-24 (2017).
14. J. Witwer, and R. Berns, "Increasing the Versatility of Digitizations through Post-Camera Flat-Fielding", in *Proceedings of IS&T Archiving Conference* (2015), pp. 110-113.
15. "Colorimetry—Part 6: CIEDE2000 Colour-Difference Formula", ISO/CIE 11664-6: 2014.
16. M. Melgosa, D. H. Alman, M. Grosman, L. Gómez-Robledo, A. Trémeau, G. Cui, P. A. García, D. Vázquez, C. Li, and M. R. Luo, "Practical demonstration of the CIEDE2000 corrections to CIELAB using a small set of sample pairs", *Color Res. Appl.* **38**, 429-436 (2013).
17. M. Melgosa, R. Huertas, and R.S. Berns, "Relative significance of the terms in the CIEDE2000 and CIE94 color-difference formulas", *J. Opt. Soc. Am. A* **21**(12), 2269-2275 (2004).
18. M. Melgosa, L. Gómez-Robledo, P. A. García, S. Morillas, C. Fernández-Maloigne, N. Richard, M. Huang, C. Li, and G. Cui, "Color-quality control using color-difference formulas: progress and problems", in *Proc. of SPIE* (2017), 104530U.
19. M. Huang, G. Cui, M. Melgosa, M. Sánchez-Marañón, C. Li, C., M. R. Luo, and H. Liu, "Power functions improving the performance of color-difference formulas", *Opt. Express* **23**, 597-610 (2015).
20. R. S. Berns, "Modification of CIEDE2000 for Assessing Color Quality of Image Archives", in *Proceedings of IS&T Archiving Conference* (2016), pp. 181-185.
21. M. Melgosa, G. Cui, C. Oleari, P. J. Pardo, M. Huang, C. Li, and M. R. Luo, "Revisiting the weighting function for lightness in the CIEDE2000 colour-difference formula", *Col. Technol.* **133**(4), 273-282 (2017).
22. R. G. Kuehni, "CIEDE2000, milestone or final answer?", *Color Res. Appl.* **27**, 126-128 (2002).
23. C. Oleari, "Color opponencies in the system of the uniform color scales of the Optical Society of America," *J. Opt. Soc. Am. A*, **21**(5), 677-682 (2004).
24. R. Huertas, M. Melgosa, and C. Oleari, "Performance of a color-difference formula based on OSA-UCS space using small-medium color differences," *J. Opt. Soc. Am. A* **23**(9), 2077-2084 (2006).
25. C. S. McCamy, H. Marcus, and J. G. Davidson, "A Color Rendition Chart," *J. Appl. Phot. Eng.* **11**(3), 95-99 (1976).
26. K. Borrino, F.V. de Azambuja, N. Sampat, and J.A.S. Viggiano, "Sensitivity analysis applied to ISO recommended camera color calibration methods to determine how much of an advantage, if any, does spectral characterization of the camera offer over the chart-based approach", in *Electronic Imaging, Digital Photography and Mobile Imaging XIII* (2017), pp. 32-36.
27. M.P. Lanaro, A. Pezzotta, M. Pessotto, and A. Rizzi, "Fidarsi o no del ColorChecker?", in *Colore e Colorimetria - Contributi Multidisciplinari* (2018), **XIVB**, pp. 13-24.
28. "Photography - Psychophysical experimental methods for estimating image quality," ISO 20462:2005.
29. H. A. David, *The Method of Paired Comparisons*, 2nd ed. (Oxford University Press, 1988).
30. T.M. Rice and T.W. Faulkner, "The use of photographic space in the development of the disc photographic system," *J. Appl. Photogr. Eng.*, **9**, 52-57 (1983).
31. D.F. Zwick, "Psychometric scaling of terms used in category scales of image quality attributes", in *Proc. Symposium of Photogr. and Electr. Image Quality* (1984), pp. 46-55.
32. A. M. van Dijk, J.-B. Martens, and A. B. Watson, "Quality assessment of coded images using numerical category scaling", in *Proc. of SPIE* (1995), **2451**, pp. 90-101.
33. B. G. Tabachnick, and L. S. Fidell, *Using multivariate statistics*, 6th ed. (Pearson Education, 2013).
34. "P.910: Subjective video quality assessment methods for multimedia applications", International Telecommunication Union (2008), <https://www.itu.int/rec/T-REC-P.910-200804-I>.
35. J. E. Gibson, M. D. Fairchild, and S. L. Wright, "Colorimetric tolerances of various digital image displays", in *Proceedings of IS&T and SID Eighth Color Imaging Conference* (2000), pp. 295-300.
36. H. Liu, M. Huang, G. Cui, M. R. Luo, and M. Melgosa, "Color-difference evaluation for digital images using a categorical judgment method", *J. Opt. Soc. Am. A* **30**(4), 616-626 (2013).
37. S. Katsigiannis, J. N. Scovell, N. Ramzan, L. Janowski, P. Corriveau, M. Saad, and G. Van Wallendael, "Interpreting MOS scores, when can users see a difference? Understanding user experience differences for photo quality," *Quality and User Experience* **3**(1), 6 (2018).
38. G. W. Snedecor, and W.G. Cochran, *Statistical Methods*, 8th ed. (Iowa State University Press, 1989).
39. D. C. Montgomery, and G. C. Runger, *Applied Statistics and Probability for Engineers*, 2nd ed. (JohnWiley & Sons, 1999).
40. A. Rizzi, "Colour after colorimetry", *Col. Technol.* **137**(1), 22-28 (2021).

Appendix 1

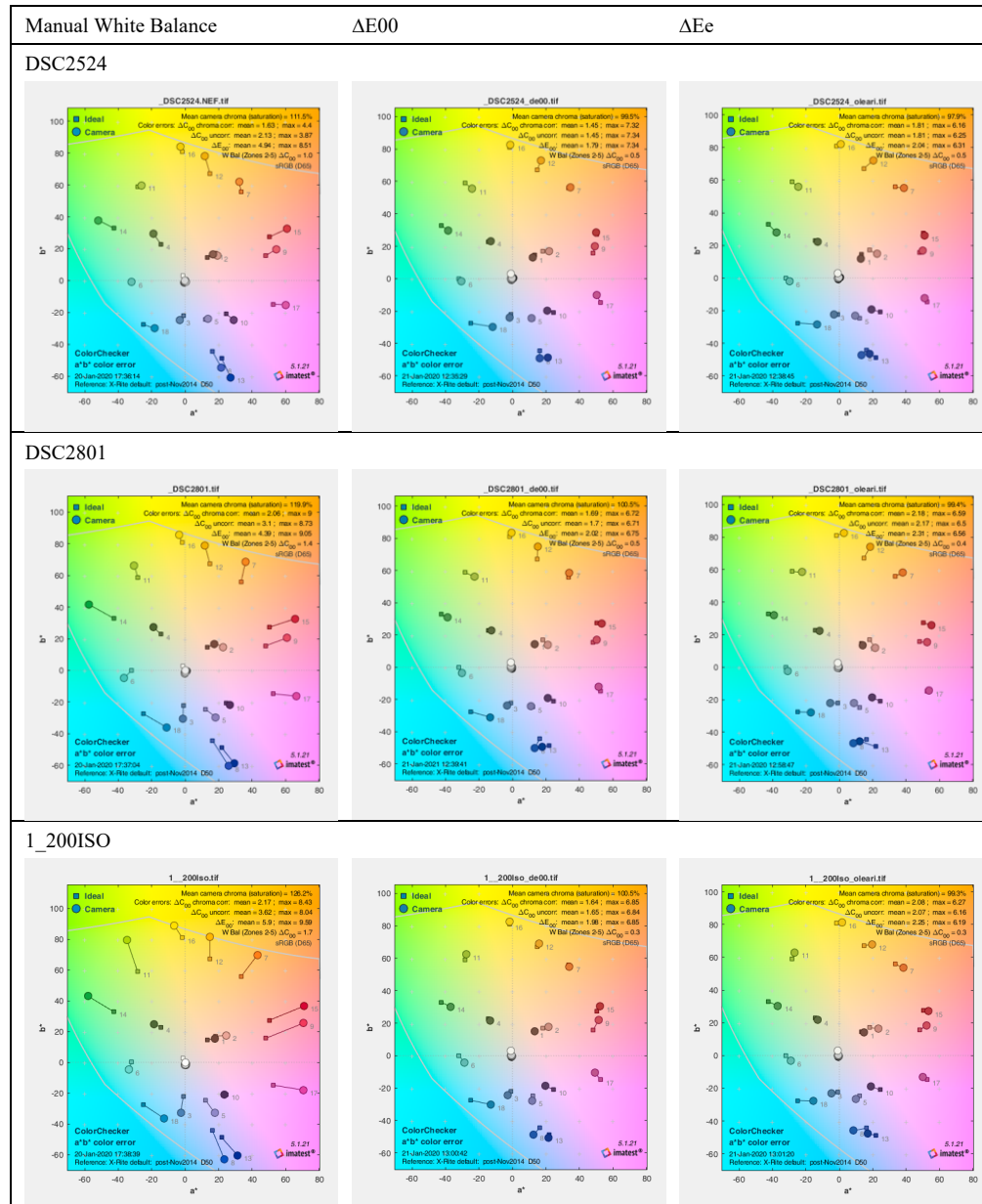
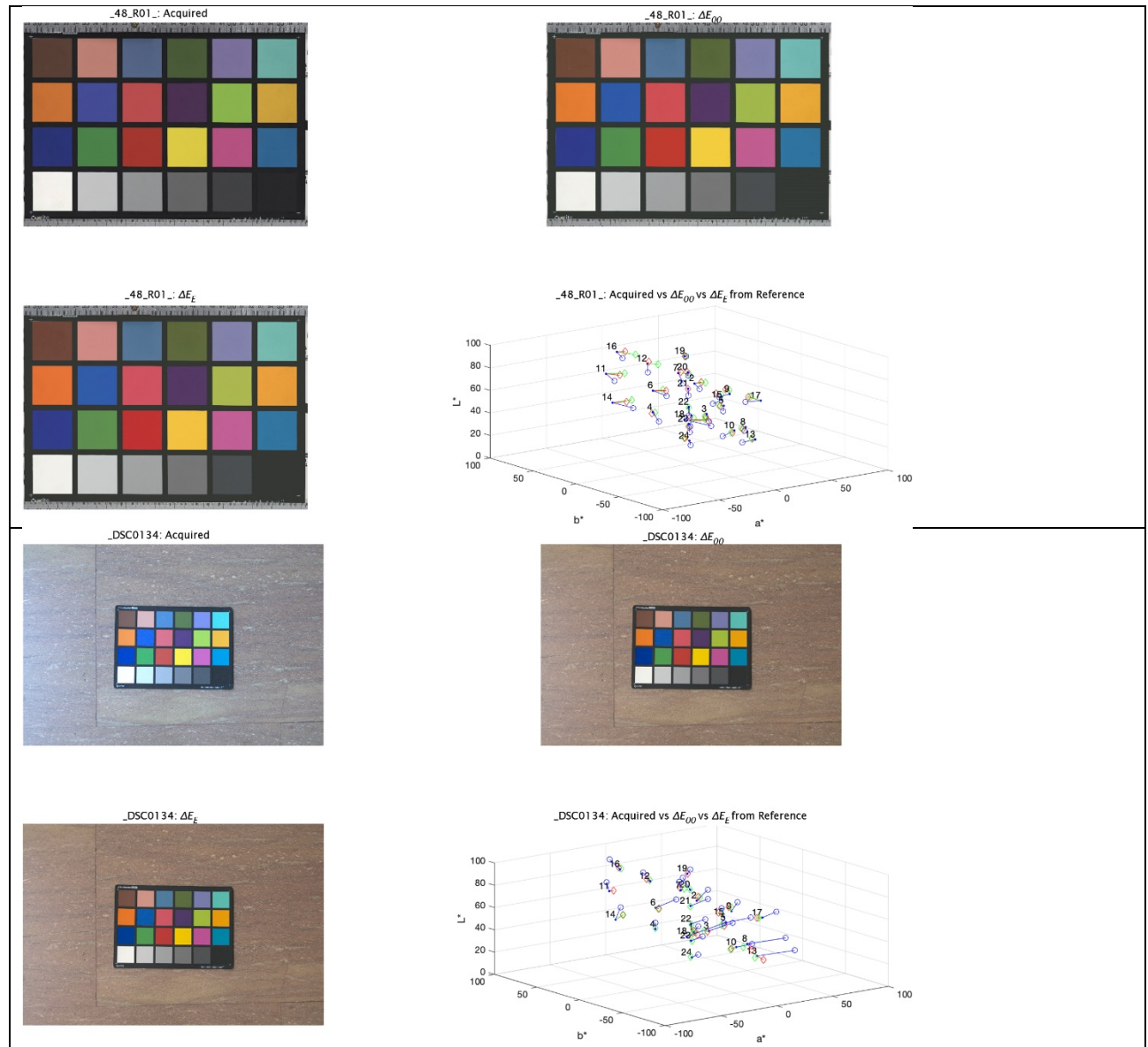


Fig. Appendix 1. Three images developed through different methods. The first column represents the result of a simple white balance using the Von Kries adaptation. In the second and the third images have been processed through the entire SHAFT development process. The second column the CC is based on ΔE_{00} in the third column the CC is based on the ΔE_c .

Appendix 2



_DSC0136: Acquired



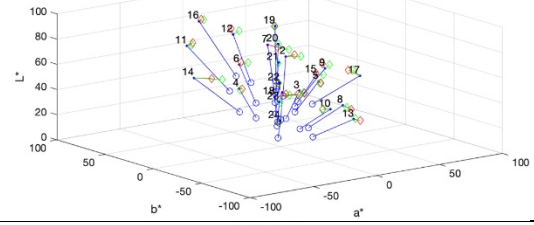
DSC0136: ΔE{00}



_DSC0136: ΔE_E



DSC0136: Acquired vs ΔE{00} vs ΔE_E from Reference



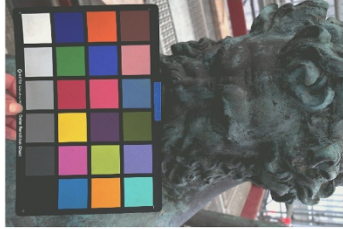
_DSC0945: Acquired



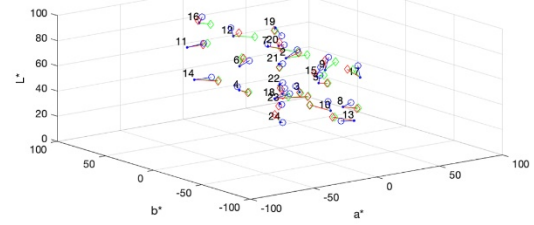
DSC0945: ΔE{00}



_DSC0945: ΔE_E



DSC0945: Acquired vs ΔE{00} vs ΔE_E from Reference



_DSC2118: Acquired



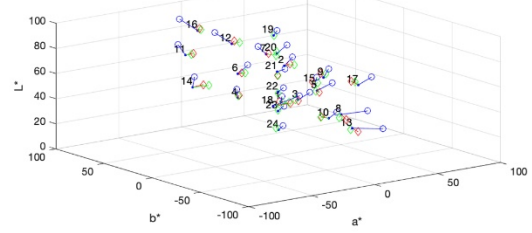
DSC2118: ΔE{00}



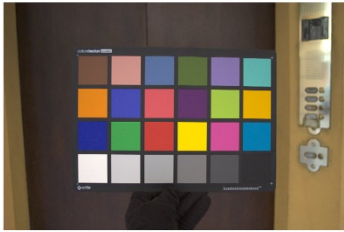
_DSC2118: ΔE_c



DSC2118: Acquired vs ΔE{00} vs ΔE_c from Reference



_DSC2416: Acquired



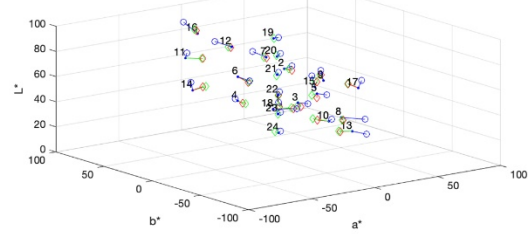
DSC2416: ΔE{00}



_DSC2416: ΔE_c



DSC2416: Acquired vs ΔE{00} vs ΔE_c from Reference



_DSC2524: Acquired



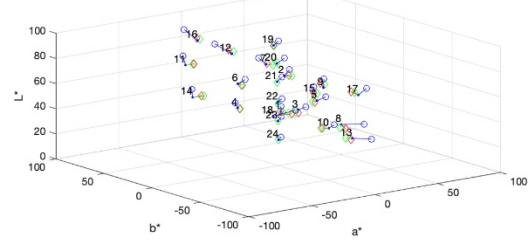
DSC2524: ΔE{00}



_DSC2524: ΔE_E



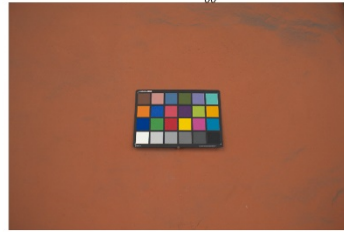
DSC2524: Acquired vs ΔE{00} vs ΔE_E from Reference



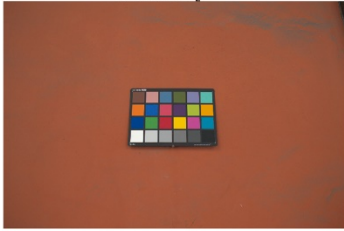
_DSC2801: Acquired



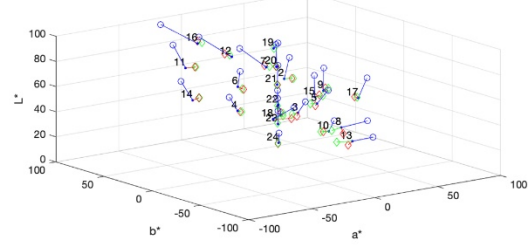
DSC2801: ΔE{00}



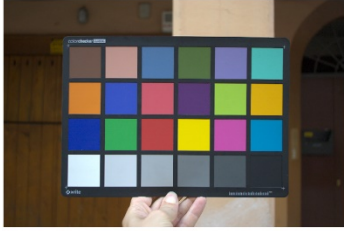
_DSC2801: ΔE_E



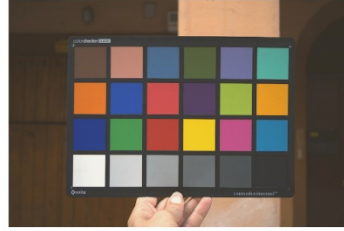
DSC2801: Acquired vs ΔE{00} vs ΔE_E from Reference



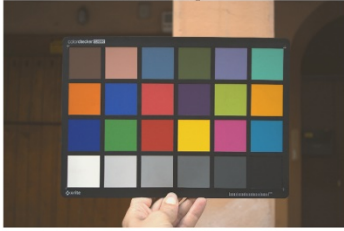
_DSC2935: Acquired



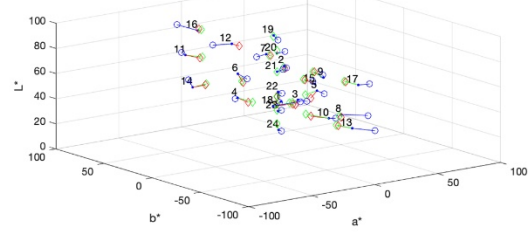
DSC2935: ΔE{00}



_DSC2935: ΔE_E



DSC2935: Acquired vs ΔE{00} vs ΔE_E from Reference



_DSC3197: Acquired



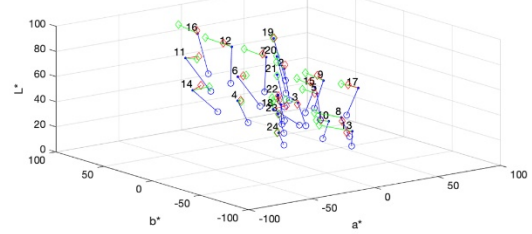
DSC3197: ΔE{00}



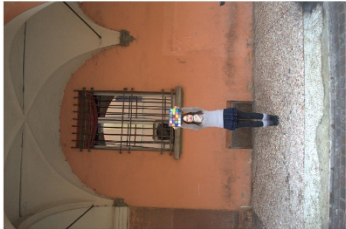
_DSC3197: ΔE_E



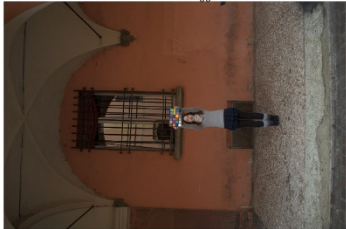
DSC3197: Acquired vs ΔE{00} vs ΔE_E from Reference



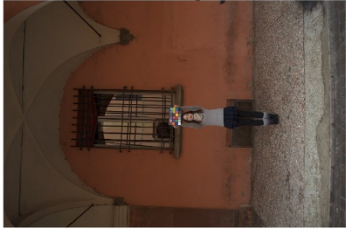
_DSC3630: Acquired



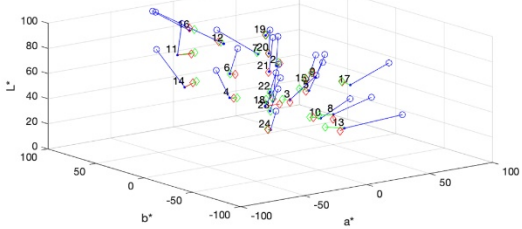
DSC3630: ΔE{00}



_DSC3630: ΔE_c



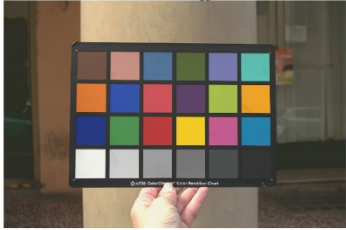
DSC3630: Acquired vs ΔE{00} vs ΔE_c from Reference



_DSC6062: Acquired



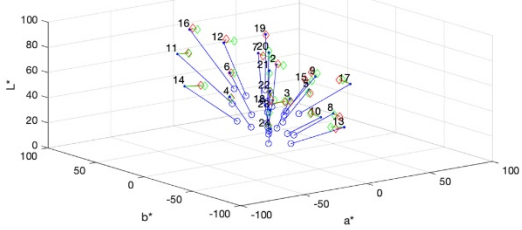
DSC6062: ΔE{00}



_DSC6062: ΔE_c



DSC6062: Acquired vs ΔE{00} vs ΔE_c from Reference



_MG_9935: Acquired



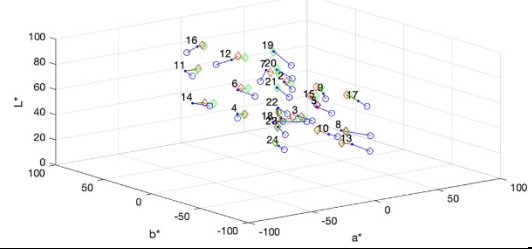
_MG_9935: ΔE_{00}



_MG_9935: ΔE_E



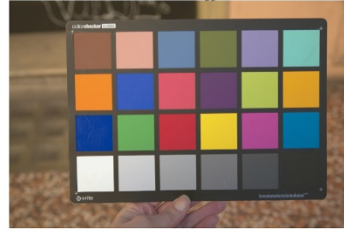
_MG_9935: Acquired vs ΔE_{00} vs ΔE_E from Reference



_T4A4626: Acquired



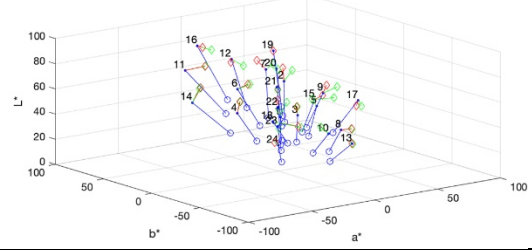
T4A4626: ΔE{00}



_T4A4626: ΔE_E



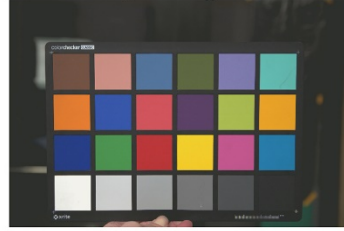
T4A4626: Acquired vs ΔE{00} vs ΔE_E from Reference



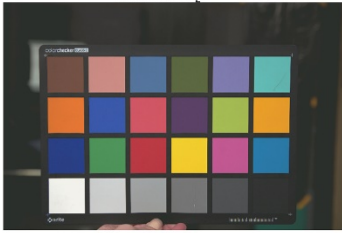
_T4A4846: Acquired



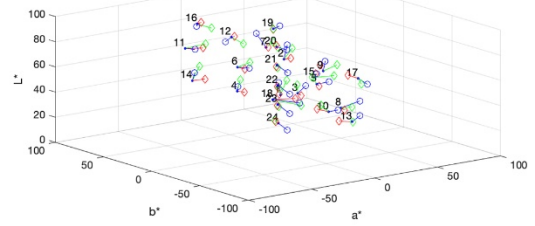
T4A4846: ΔE{00}



_T4A4846: ΔE_E



T4A4846: Acquired vs ΔE{00} vs ΔE_E from Reference



_T4A5900: Acquired



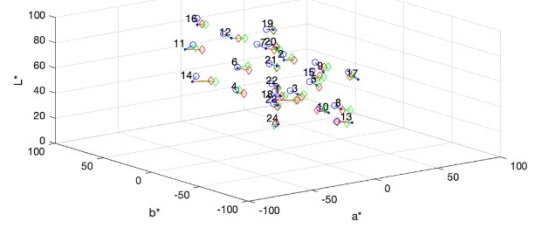
T4A5900: ΔE{00}



_T4A5900: ΔE_E



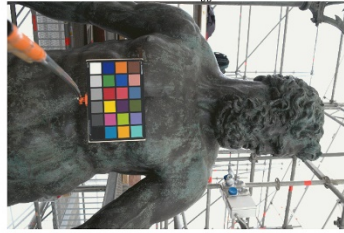
T4A5900: Acquired vs ΔE{00} vs ΔE_E from Reference



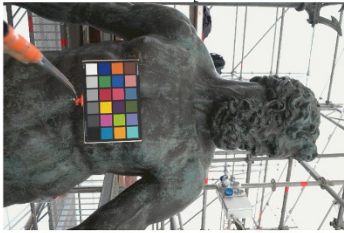
_wb_L5_A: Acquired



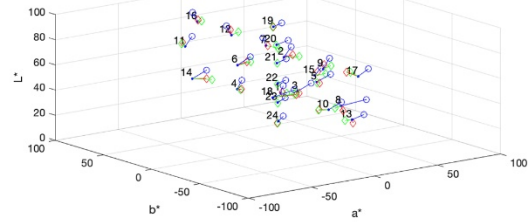
_wb_L5_A: ΔE_{00}



_wb_L5_A: ΔE_t



_wb_L5_A: Acquired vs ΔE_{00} vs ΔE_t from Reference



1__2001s: Acquired



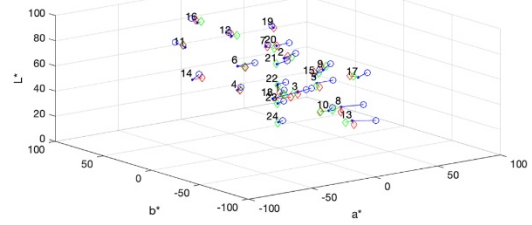
1__2001s: ΔE_{00}



1__2001s: ΔE_t



1__2001s: Acquired vs ΔE_{00} vs ΔE_t from Reference



2__2001s: Acquired



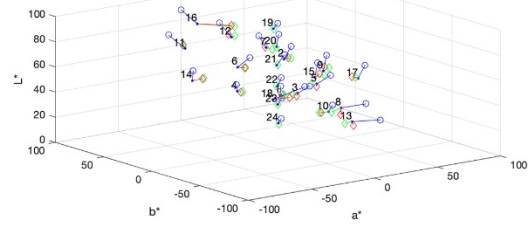
2__2001s: ΔE_{00}



2__2001s: ΔE_t



2__2001s: Acquired vs ΔE_{00} vs ΔE_t from Reference



3__2001s: Acquired



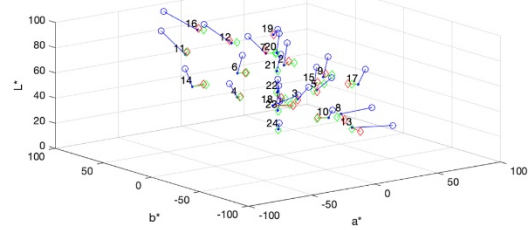
3__2001s: ΔE_{00}



3__2001s: ΔE_t



3__2001s: Acquired vs ΔE_{00} vs ΔE_t from Reference



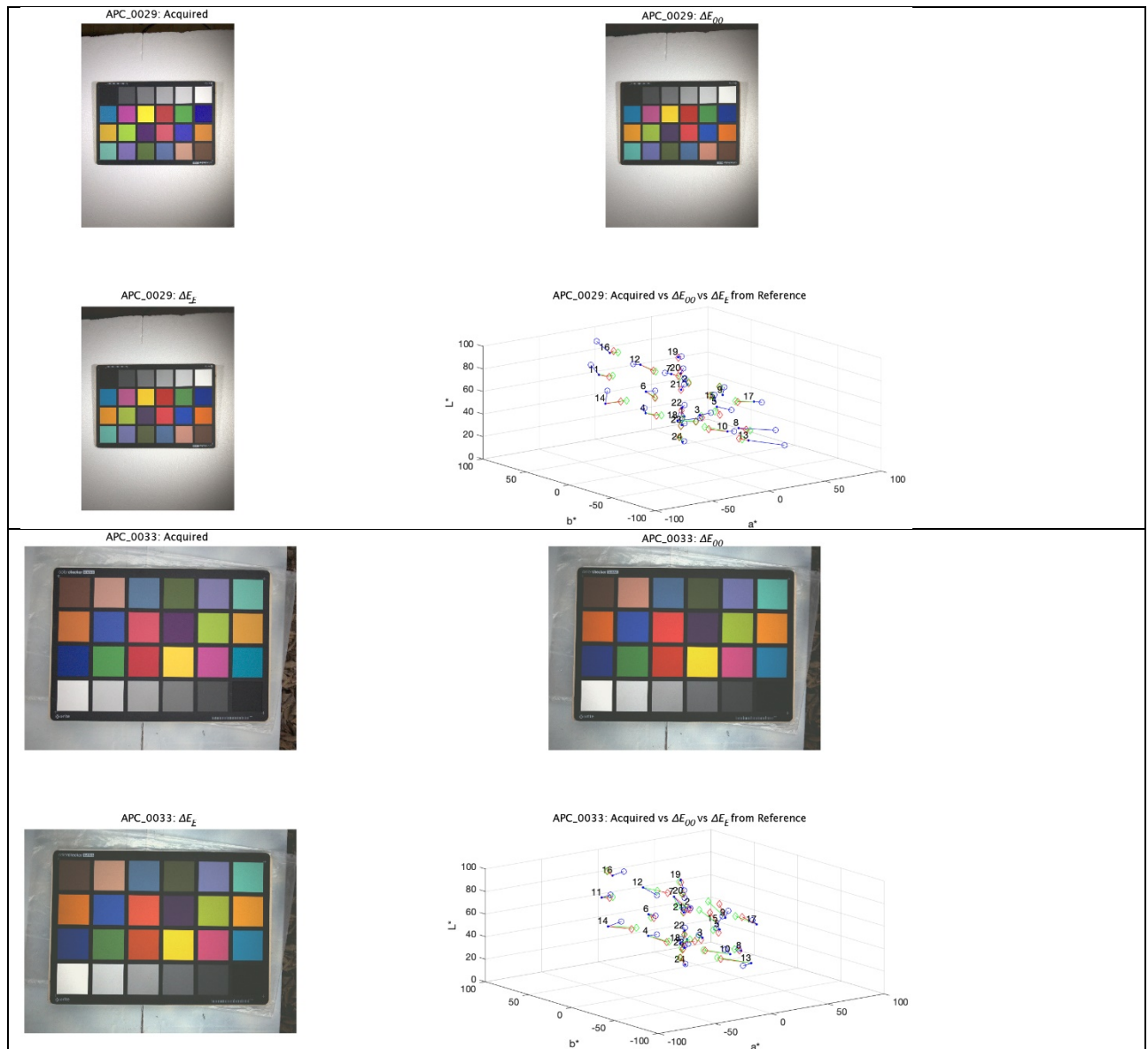
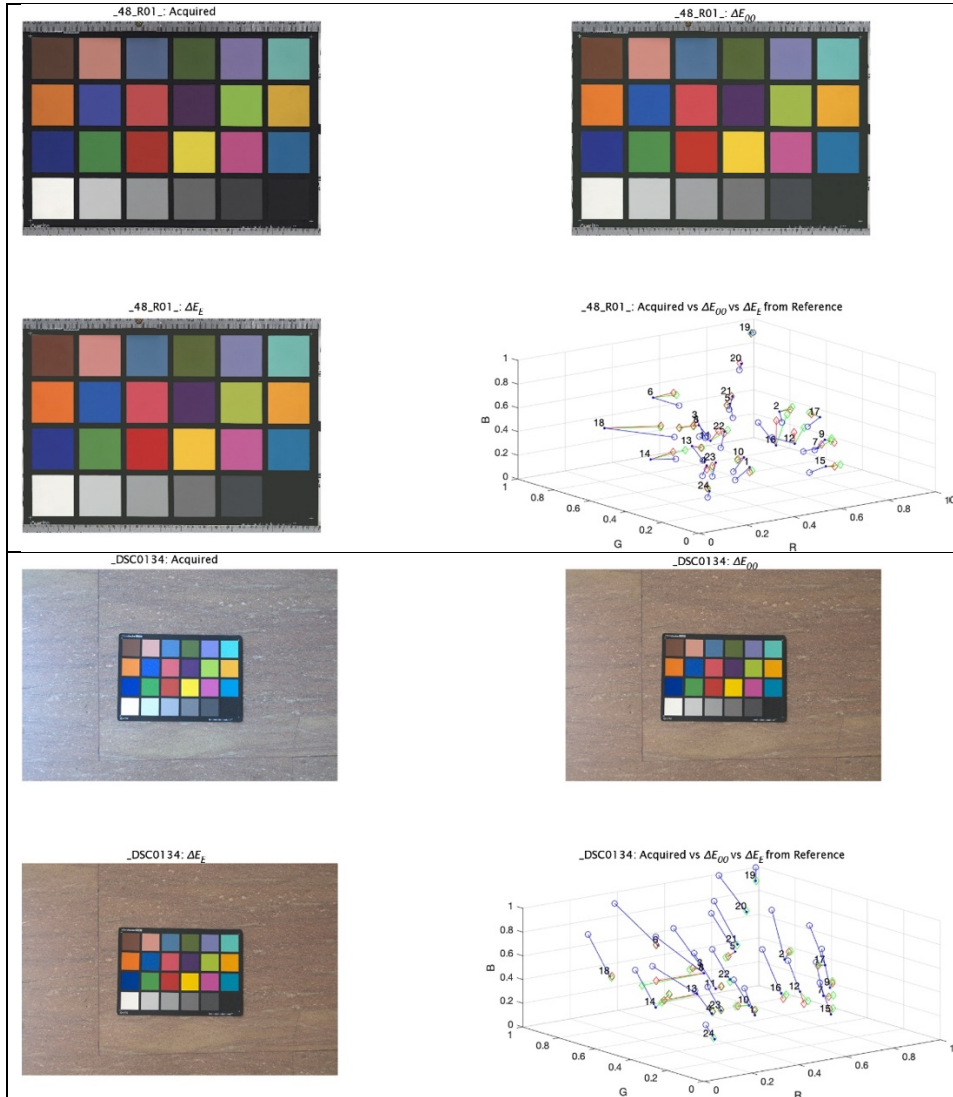
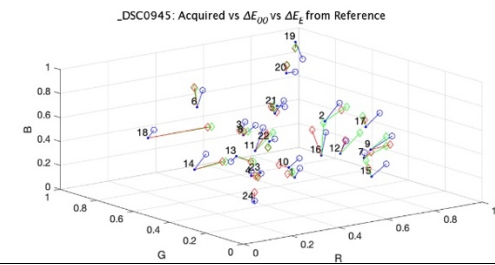
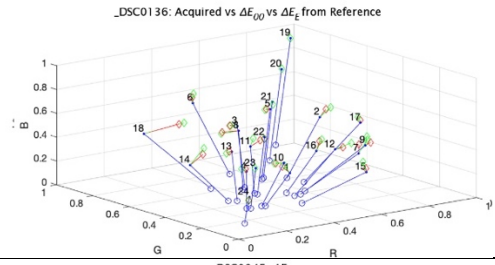
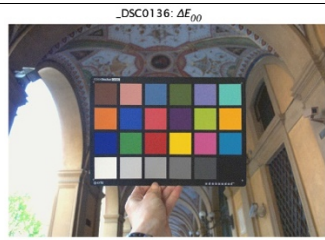
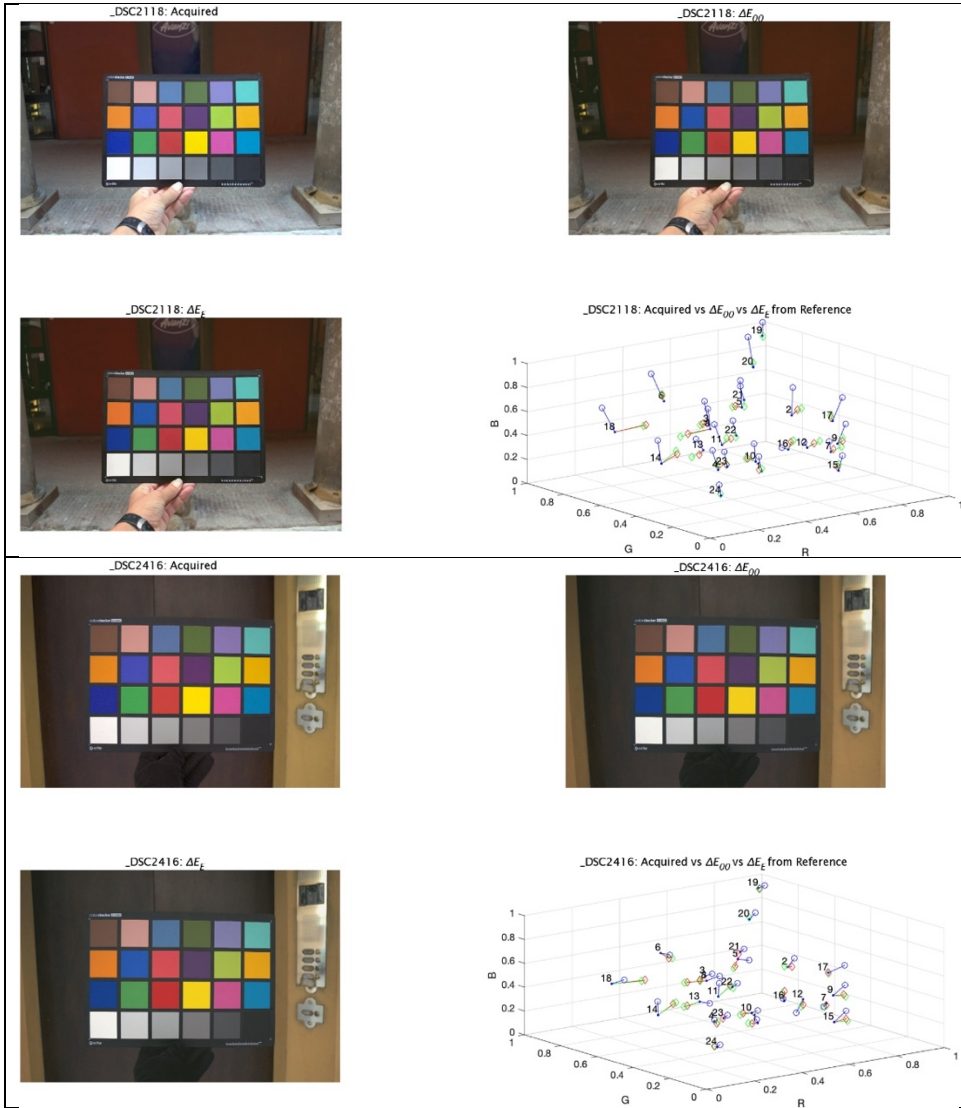


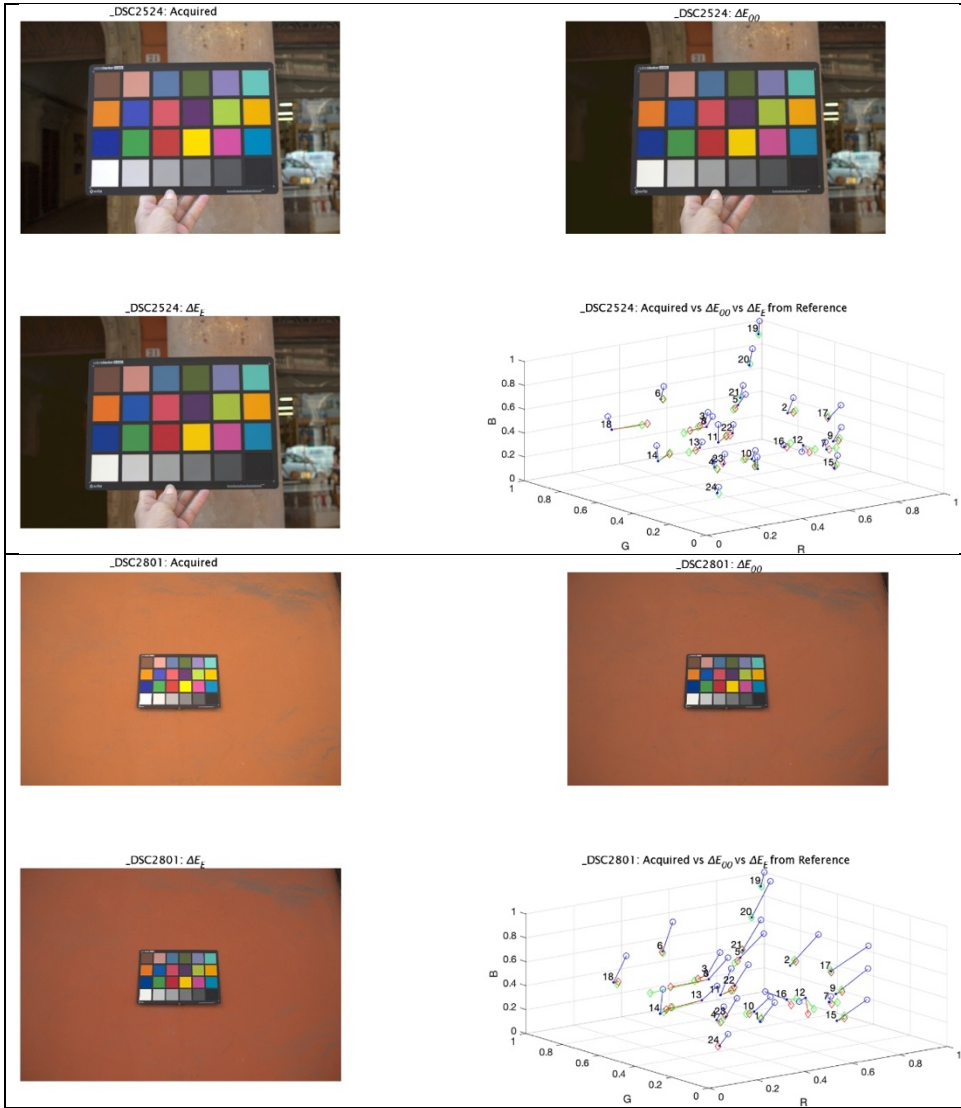
Fig. Appendix 2. Images acquired on top-right, corrected with ΔE_{00} on top-right and corrected with ΔE_E on bottom-left, respectively. On bottom-right distances from reference to acquired (blue), to ΔE_{00} corrected (red), to ΔE_E corrected (green), respectively.

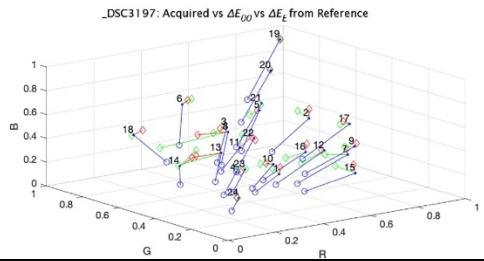
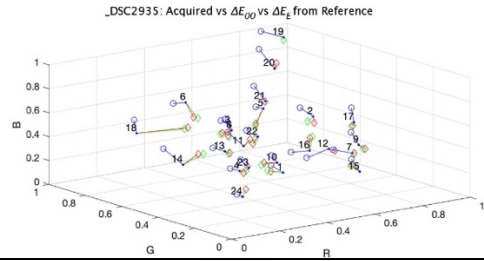
Appendix 3

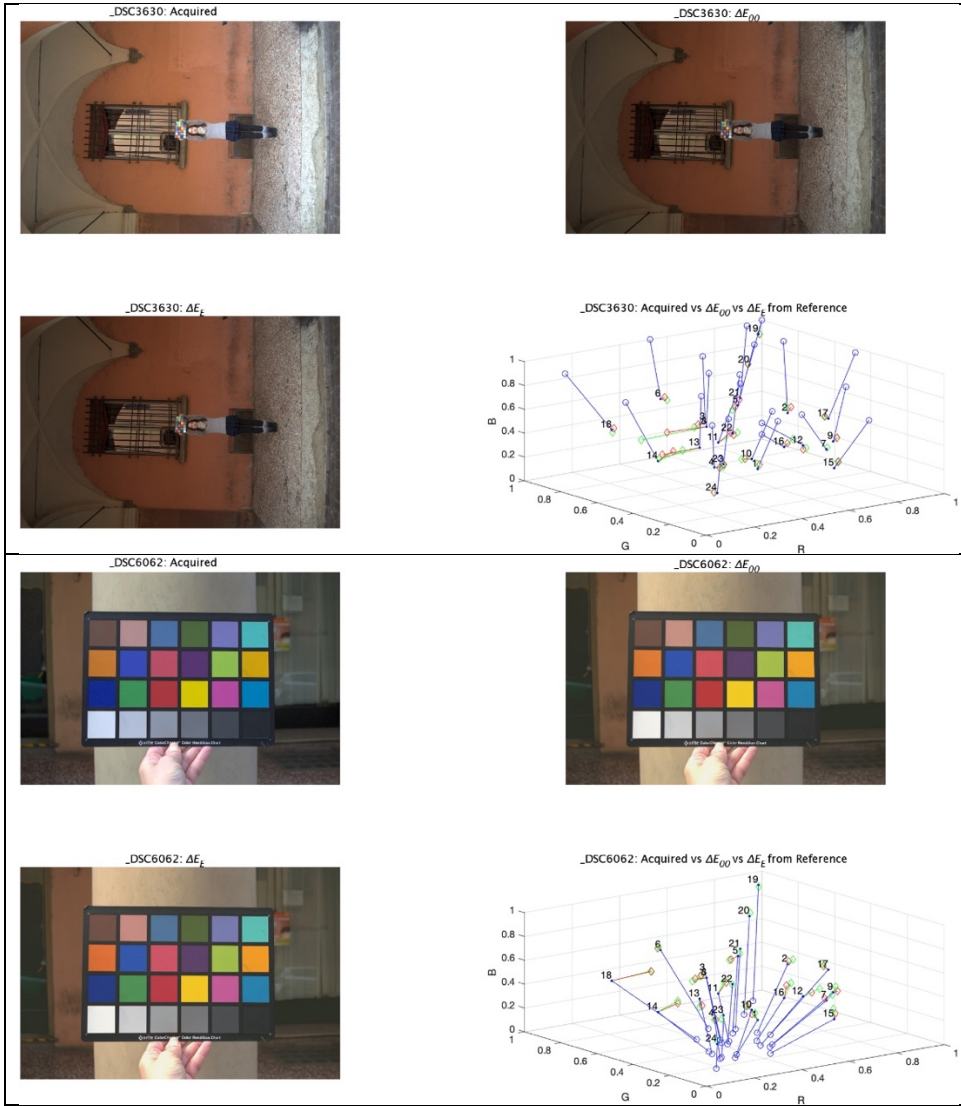


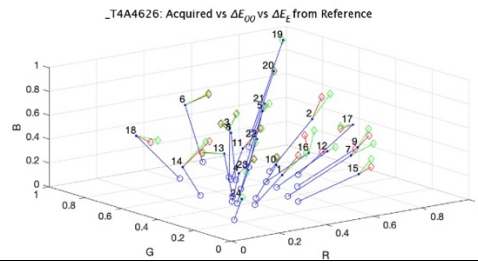
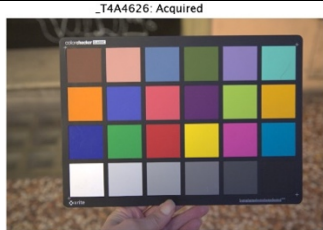
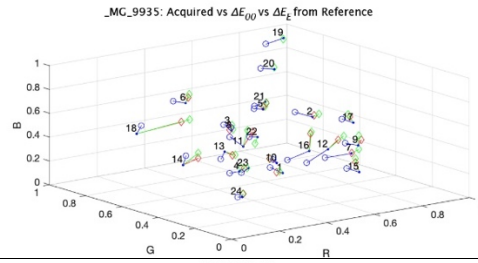


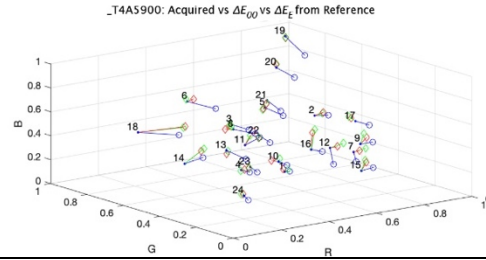
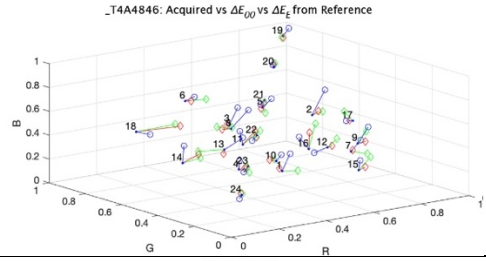
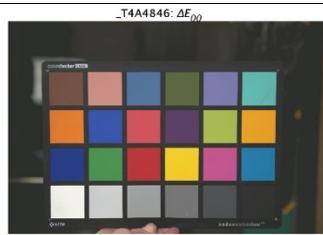


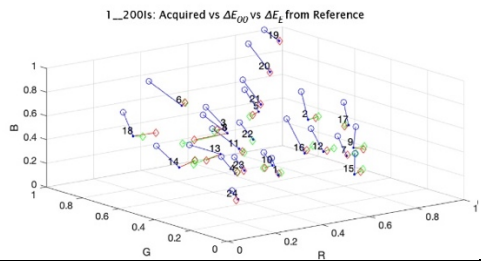
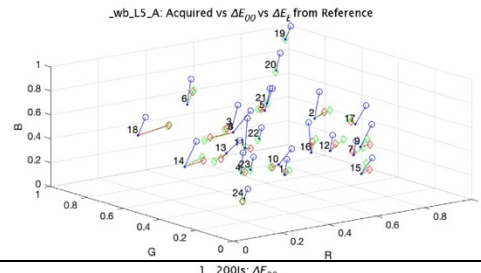
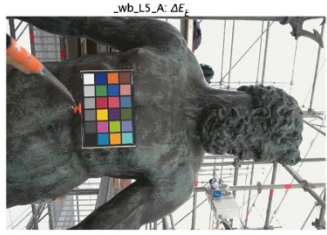
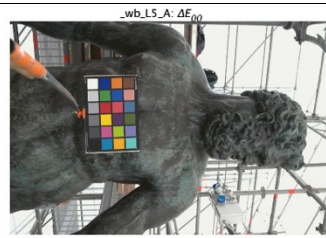
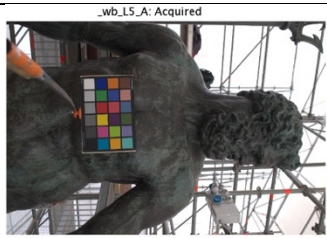


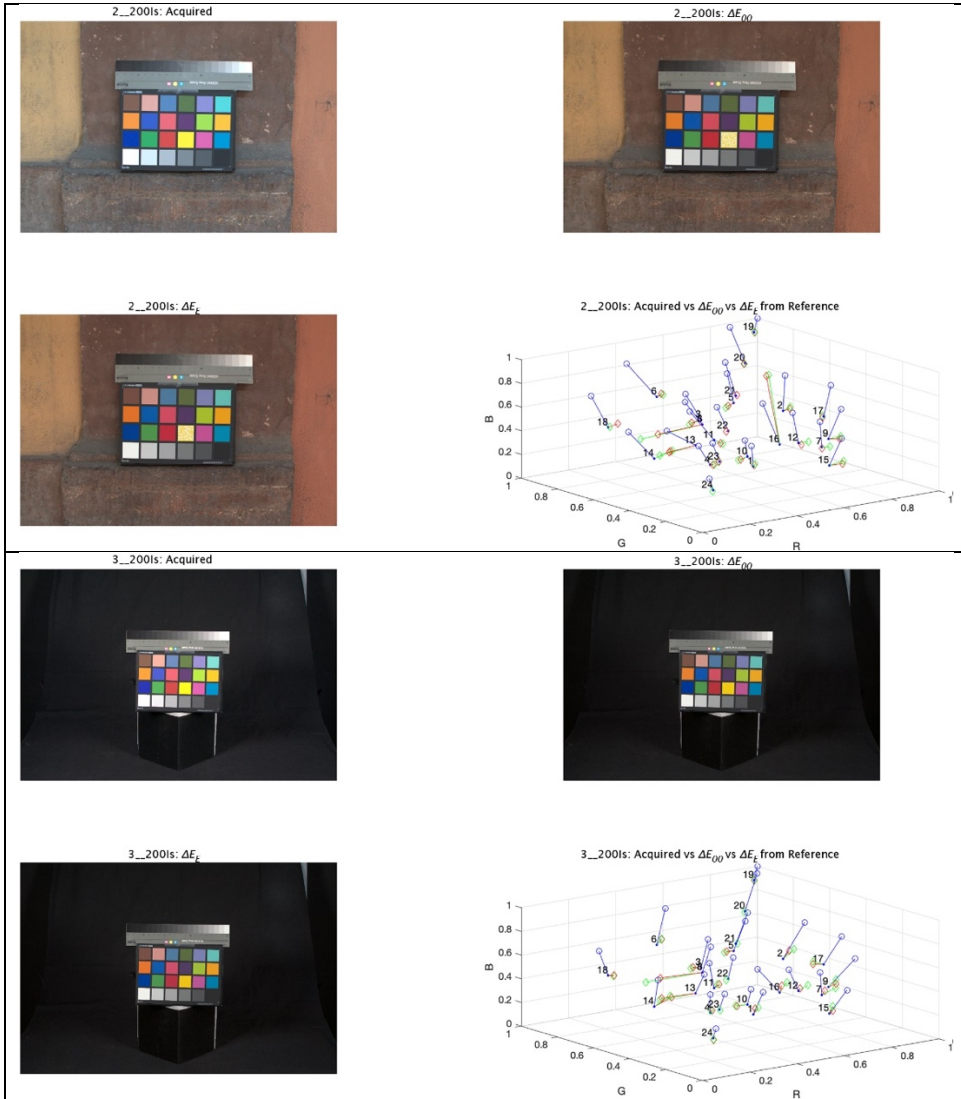












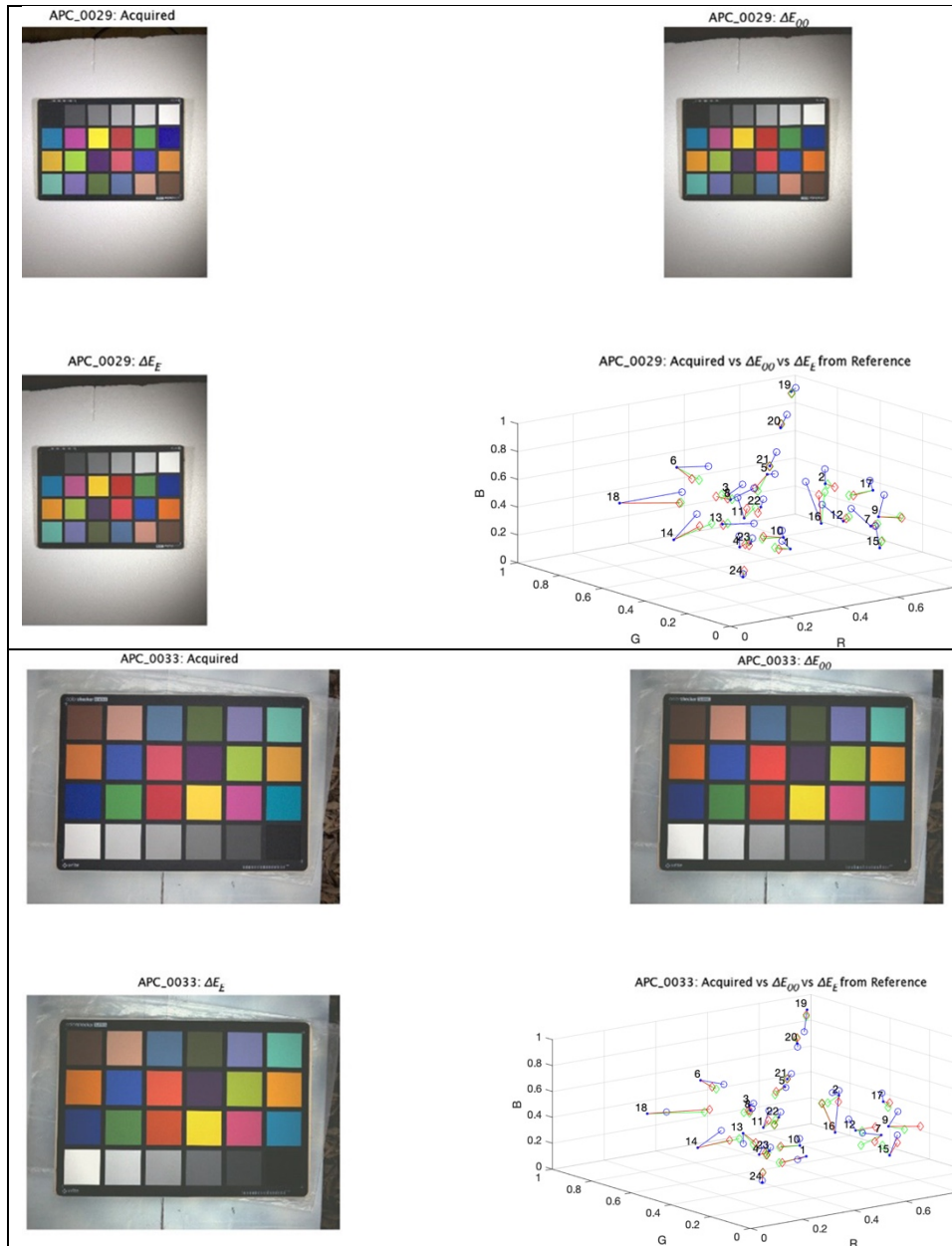


Fig. Appendix 3. Shifts expressed in raw values. On top-left images acquired; on top-right images corrected with ΔE_{00} and on bottom-left images corrected with ΔE_z . On bottom-right distances from reference to acquired (blue), to ΔE_{00} corrected (red), to ΔE_z corrected (green), respectively.