ABSTRACT
High dynamic range (HDR) display prototypes have been built and used for scientific studies for nearly a decade, and they are now on the verge of entering consumer market. However, problems exist regarding the accurate color reproduction capabilities on these displays. In this paper, we first characterize the image reproduction capability of a state-of-art HDR display through a set of measurements, and present a novel calibration method that takes into account the variation of the chrominance error over HDR display’s wide luminance range. Our proposed 3D octree forest data structure for representing and querying the calibration function successfully addresses the challenges in calibrating HDR displays: (i) high computational complexity due to nonlinear chromatic distortions; (ii) huge storage space demand for a look-up table. We show that our method achieves high color reproduction accuracy through both objective metrics and a controlled subjective study.

Index Terms— HDR, display calibration, color calibration

1. INTRODUCTION
HDR displays offer an unparalleled viewing experience due to having a peak luminance that is an order of magnitude higher than conventional displays, and similarly notably low black levels thanks to local dimming technology. Recent subjective studies show that the viewers strongly prefer the extra dynamic range offered by such displays over the conventional viewing experience [1].

The advantages of the HDR display technology have also been recognized by the industry. Professional cameras have been able to natively capture HDR, whereas content production studios are starting to slowly move towards HDR workflows (e.g. for color grading). On the other hand, standardization of HDR content distribution is an ongoing effort carried out by relevant committees in SMPTE, JPEG, MPEG, Blu-ray, with strong interests from industry representatives. Meanwhile, in addition to the prototypes from Brightside, Dolby and SIM2, new “prosumer” HDR displays are being announced.

The advent of HDR content creation workflows and distribution systems to the home promises new opportunities for content producers. Color graders can exploit the higher dynamic range for better expression of their artistic intent, and viewers in the home can enjoy higher quality content. To accurately encode the artistic intent in the distributed content and faithfully reproduce the artistic intent in the home, the use of calibrated HDR displays is indispensable, both during color grading and viewing in the home.

The color calibration of an HDR display is challenging because of two reasons: (i) the high nonlinearity in chromatic distortion in dependence of luminance, which requires high computational complexity to compensate for it if an analytic calibration function is used, (ii) the high bit depth used for representing colors for HDR displays, which practically prevents the use of an look-up table generated by brute force sampling of the calibration function. These challenges set significant obstacles in utilizing the full potential of HDR displays where precise reproduction of colors is required. In this work, we show that these challenges can be alleviated. Specifically, we make the following contributions:

• We characterize the SIM2 HDR display’s image reproduction capability through a comprehensive set of measurements.
• We propose to use of a 3D octree forest data structure to enable efficient representation and querying of an analytic luminance dependent calibration function.
• We present a complete pipeline for HDR display calibration that takes into account the variation of the chrominance error over the luminance range. We evaluate our method objectively, and also through a controlled subjective study.

2. BACKGROUND
Calibration of Standard Dynamic Range (SDR) displays is less challenging than the calibration of HDR displays, since the color reproduction capability of SDR display devices does not vary as much as it does in HDR displays. In practice, often an ad-hoc calibration by applying some affine transform that compensates for the errors in the white point and the primaries of the color gamut can be sufficient for SDR displays. For a comprehensive review of the calibration of projection systems we refer the reader to Bimber et al. [2]. Examples of recent work include [3–5].

HDR displays have been introduced nearly a decade ago [6], and since then multiple research prototypes have been presented. Extensive discussions exist for the evolution of HDR imaging and display technologies [7]. The custom-built HDR display prototypes [8–12] have often been designed and used for experimental purposes such as performing various perception studies. Other, more general purpose HDR displays have been introduced by private enterprises such as Brightside, SIM2 and Dolby. These displays have also been extensively utilized in the research community for various perception and preference experiments [1, 13–18].

Calibration of HDR displays has received little attention in the research community so far. Seetzen et al. discussed a calibration [19] procedure specific to the dual modulation technology where they aim at accounting for the chromaciticy variations in LED backlights. Ruppertsberg et al. [20] presents a characterization of a Brightside DR37-P HDR display and provide a simple affine mapping from input RGB values to the displayed luminance and chrominance. As we discuss later, such an affine mapping is not well suited for compensating the non-uniform chrominance errors.

Color gamut mapping is a related, but different research problem. In standard gamut mapping, an input gamut is mapped to a...
(possibly smaller) previously known output gamut. In HDR display calibration the output gamut is not known a priori and has to be measured first. Moreover, the output color gamut depends strongly on the luminance level. As such, the mapping function required for HDR display calibration has additional layers of complexity compared to gamut mapping in the classical sense. For a comprehensive background in gamut mapping, we refer the reader to Morovic et al. [21] and Banterle et al. [22].

3. DISPLAY ANALYSIS

In this section we analyze the color reproduction properties of a SIM2 HDR display [23], which in the remainder of this paper we refer to as simply the HDR display. The HDR display utilizes the dual modulation technology and comprises a full HD LCD front panel and an LED backlight matrix consisting of 2202 individually controllable LED elements. In the HDR mode, the native input to the display is visual content encoded in LogY, u’, v’ format, where logY represents log luminance, and u’ and v’ are the chromaticity values according to the CIE 1976 u’-v’ chromaticity diagram [24]. The peak luminance of the HDR display is 4000 cd/m².

The HDR display, like other current consumer high definition displays, conforms to the Rec. 709 [25] standard for color reproduction. Rec. 709 standard is currently used for professional high definition content production and distribution, and specifies image format parameters such as the primary colors and the reference white point. In the remainder of this section we present our findings from two sets of measurements with respect to Rec. 709 input:

M1 We measured reproduced colors, given the reference white and the three primary colors specified by Rec. 709 standard, with different luminances as input to the HDR display.

M2 We measured reproduced colors, given densely sampled input colors from an area that encompasses the Rec. 709 gamut, with different luminances as input to the HDR display.

Our key observation from M1 is the dependence of the displayed primary colors to the input lumiance. This is illustrated in Figure 1, which e.g. shows that u’-v’ coordinates of the reproduced primary colors vary strongly at three example luminance levels, 0.03, 48, and 1916.5 cd/m². Based on our measurements at these and other luminance levels, we conclude that all reproduced primary colors tend to loose saturation below 0.08 cd/m² or above 2000 cd/m², while the reproduced blue primary color looses saturation rapidly above 125 cd/m².

\[ G_i \mapsto \chi(\chi^{-1}(G_i)) \]

\[ D : \chi^{-1}(G_i) \rightarrow G_o, \quad c_i \mapsto c_o \]

\[ \chi(c_i) = \chi(D(c_i)), \quad c_i \in G_i \]

where \( \chi \) maps a trichromatic color to its chromaticity representation according to the u’v’ chromaticity diagram [24]. In practice, a real physical display never perfectly fulfills this property. Usually the output gamut volume is smaller, i.e. \( G_o \subset G_i \) like for our HDR display. Therefore, it is highly desirable that the display has an accurate chromaticity reproduction at least for input colors which are also in the output gamut volume available. Thus, our first goal is to develop a technique to capture the chromaticity coordinates obtained in M2. In Figure 2, we observe strong chromaticity errors within the same luminance level, as well as a variation in chromaticity errors at different luminance levels. These measurements suggest that a global linear transform is most likely not sufficient to account the HDR display’s chromaticity errors.

Despite the nonlinearity being an obstacle for a global transformation solution, the error vector field changes mostly smoothly in both luminance and chromaticity directions. Equipped with the aforementioned observations, we next discuss our calibration method.

4. CALIBRATION METHOD

We define our goal as minimizing the chromaticity error of the HDR display, i.e. reducing the differences between input and reproduced chromaticities as much as possible. To that end, we use sparse measurements of the HDR display’s 3D gamut volume and exploit the smoothness assumption of the error vector field to develop an accurate and efficient chromaticity calibration method for HDR displays. We model the HDR display’s color reproduction by a mapping

\[ D : G_i \rightarrow G_o, \quad c_i \mapsto c_o \]

which receives input colors \( c_i \) as input, and reproduces output colors \( c_o \). The set of all valid input colors \( G_i \) creates the input gamut volume, while the set \( G_o = D(G_i) \) creates the output gamut volume. The input gamut volume of the HDR display consists of all colors of the Rec.709 gamut in a luminance range up to 4000 cd/m². Figure 1 indicates that it contains the output gamut volume. A perfect chromatically calibrated HDR display has the property

\[ \chi(c_i) = \chi(D(c_i)), \quad c_i \in G_i \]

1See our video [26] for additional visualizations.
a calibration function $K : \mathcal{G}_0 \rightarrow \mathcal{G}_i$ such that $\chi(e_i) \approx \chi(D(K(e_i)))$ for all $e_i \in \mathcal{G}_i$. This essentially means that $K$ has to approximate the inverse of the display’s color reproduction function as close as possible, i.e., $K \approx D^{-1}$. Based on our observations in Section 3, we further request that $K$ should describe a smooth error field [also illustrated in Figure 2], i.e., it should be a smooth mapping.

**Algorithm 1 Octree construction.**

1: procedure INIT($C_i, S, d_{max}$)
2: \[\text{if } C \cap S \neq \emptyset \text{ and } d_{max} > 0 \text{ then} \]
3: \quad subdivide $C$ in 8 equal sub-cells $C_1, \ldots, C_8$
4: \quad for each $i$ from 1 to 8 do
5: \quad \quad INIT($C_i, S, d_{max} - 1$)
6: \quad end for
7: \text{else} \]
8: \quad for each $i$ from 1 to 8 do
9: \quad \quad if $p_i^C \in \mathcal{G}_0$ then
10: \quad \quad \quad store $q_i^C := K_{tps}(p_i^C)$ at vertex $v_i^C$
11: \quad \quad end if
12: \quad end for
13: end if
14: end procedure

Figure 3 illustrates the major steps of our calibration method. In this pipeline the raw measurements are first cleaned up such that we end up with one-to-one correspondences between the output and the input colors. The resulting sparse correspondences are utilized for dense interpolation with the thin-plate spline (TPS) method [27] giving an analytic expression for $K$. In the last step, the interpolated function $K$ is adaptively sampled using an octree forest data structure to obtain a storage and run-time efficient representation of the calibration function. Next, we discuss these steps in detail.

### 4. Measurement and interpolation

We conducted measurements across a wide luminance range and gamut using a Photo Research Spectrascan PR730 spectrometer. During each measurement, a color patch is displayed in the center of the HDR display while the spectrometer records the measured Y, $u'$, $v'$ color values. To measure colors which are uniformly distributed across the whole output gamut volume of the display, we sample the input gamut volume $\mathcal{G}_i$ at concentric triangular prisms that are similar to the prism of the Rec. 709 gamut (Fig. 2a) and remove measurements from the set which contribute to a non-uniform distribution. To achieve this, we sequentially compute for each measured color $c_o$ error vectors $e := c_o - c_i$ and $\bar{e} := \bar{c}_o - c_i$, where $\bar{c}_o$ is the average of the 4 closest measured colors to the current color $c_o$, and remove measurements from the set with an angle $\angle(e, \bar{e}) > 35^\circ$.

We derive an approximation of the inverse of the display’s color reproduction function $D^{-1}$ by using a smooth interpolant which passes through all colors of the cleaned up set of measured colors. We employ a thin-plate spline (TPS) [27] based scattered data interpolation approach to derive $K_{tps}$, because TPS guarantees a globally smooth interpolation between samples and it is shown [4, 5] that it is well suited for color space interpolation.

The run-time required for computing a single calibrated input color with the TPS method scales linearly with the number of measurements $k$ that were used for the derivation of $K_{tps}$ [5]. In our experiments we used $k = 975$ measurements in order to nicely cover the output gamut volume. Instead of computing the calibrated input color for each image pixel, we can also sample $K_{tps}$ at all possible uncalibrated input colors and store the computed calibrated output colors in an indexed look-up table (LUT), which then allows to query the function in constant run-time. For the HDR display, which represents input colors with 33 bits (luminance 13 bits, chrominance 20 bits), such a LUT would have a size of $2^{33} \cdot 33$ bits $\approx 35.5$ GB. In comparison, a traditional SDR display with 24 bits colors requires only 50 MB for a LUT. To enable a controlled trade-off between (i) computation required for querying a function value and (ii) memory consumption, we propose the use of an octree forest data structure which is described in the following section.

### 4.2. Octree forest generation and query

To generate an octree forest, first a regular 3D grid is created by subdividing the bounding box of the output gamut volume $\mathcal{G}_i$ into $M \times N \times L$ grid cells where all edges are aligned along $u'$, $v'$, or logY direction. All edges in the $u'$-$v'$-plane have equal length, while the length of edges in logY direction is also constant to approximately reflect the luminance sensitivity of the human visual system. In each grid cell we construct octrees which can adaptively sample $K_{tps}$. In a grid cell $C$, which has vertices $v_1^C, \ldots, v_8^C$ located at positions $p_1^C, \ldots, p_8^C$, we construct an octree according to Algorithm 1. It uses as input a set of guidance colors $S \subset \mathcal{G}_o$, which are used to guide the octree construction algorithm. Grid cells are recursively subdivided into sub-cells until no colors from $S$ are located in a leaf sub-cell or the maximal depth $d_{max}$ is reached. At the vertices of the smallest sub-cells, sampled values of $K_{tps}$ are stored to be later used for interpolating arbitrary function values of $K_{tps}$. Examples of guidance colors are colors from the skin gamut area, from the area around the reference white point, or from the area close to the boundary of $\mathcal{G}_o$ to increase the coverage of this region with sub-cells completely contained in $\mathcal{G}_o$. Figure 3 shows an example $u'$-$v'$-slice of an octree forest constructed from guidance colors from the vicinity of the boundary of $\mathcal{G}_o$. In our experiments, we create a 3D grid consisting of $20 \times 20 \times 10$ cells, where cells close to the boundary of $\mathcal{G}_o$ are further subdivided by octrees of a maximal depth of 4 (Figure 4). The whole data structure consumes around 100kB of storage space.

To query a value $K_{tps}(c_i)$ for an arbitrary uncalibrated color $c_i$, we first identify the largest cell containing this color. This is achieved in constant run-time due to the regular grid structure. The octree within this cell is then traversed up to the smallest sub-cell $C_{sub}$ containing $c_i$. $K_{tps}(c_i)$ is then approximated by trilinear interpolation [28] based on the uncalibrated colors $p_1^{C_{sub}}, \ldots, p_8^{C_{sub}}$ and corresponding calibrated colors $K_{tps}(p_1^{C_{sub}}), \ldots, K_{tps}(p_8^{C_{sub}})$ stored in the corner vertices of $C_{sub}$.

### 5. EVALUATION

In this section, we evaluate our calibration method, both objectively through numerical error analysis in the $u'$-$v'$ space, and subjectively.
5.1. Objective Evaluation

To obtain a numerical estimate of our calibration method’s performance, we first displayed the colors of the Macbeth color chart on our calibrated HDR display both using an affine transform method and the 3D octree forest method, as well as without any calibration. We computed the affine transform by minimizing the Euclidean distance between input colors and measured output colors in the least-squares sense. The color chart comprises patches which have spectral reflectances intended to mimic those of natural objects such as human skin, foliage, and flowers. Note that during calibration, we specifically did not sample these colors. We measured the \( u' \) and \( v' \) values of the displayed colors for both calibrations and the uncalibrated case. Equipped with this data, we computed the errors in all three cases with respect to the actual \( u' \) and \( v' \) values of the Macbeth colors. Figure 5 shows the magnitude of deviation in the \( u'v' \)-space from the ideal Macbeth colors.

The error bars in Figure 5 indicate that both the affine transform and the 3D octree forest methods notably improve the color reproduction accuracy of the HDR display. We also note that this improvement is more pronounced for the 3D octree method. To verify this improvement also in subjective terms, we present the results of a subjective study in the next section.

5.2. Subjective Evaluation

We performed a controlled two-alternative-forced-choice study of pairwise comparisons in order to assess if our method’s advantage in objective terms (Section 5.1) also translates to the subjective user experience. Our experiment setup comprised a calibrated 10-bit Eizo ColorEdge CG277 display (reference), and a SIM2 HDR47ES4K display (test) placed side by side in a dimly illuminated room. Our stimuli consisted of 5 natural and 3 synthetic images. At each trial, the stimulus image was displayed on the reference display. On the test display, the subjects had the ability to view the same image either with or without calibration and they were able to switch back and forth between these two versions. The subjects were asked to compare both the calibrated and native versions shown on the test display with the same image shown on the reference display, and assess which one has more similar colors with respect to the reference.

The experiment was preceded by a short training session. During actual experiment, each subject evaluated 16 pairs that consisted of 8 images displayed at two luminance ranges: 0.09 – 110 cd/m\(^2\) (standard) and 0.06 – 1454 cd/m\(^2\) (high). The maximum luminance of the reference display was set accordingly to 110 cd/m\(^2\). While we did not set any time limitations, our 10 subjects finished the task in approximately 5 minutes on average.

In our study, 93.75% of the pairwise comparisons resulted in favor of our calibration method over the non-calibrated version, indicating that our calibration method brings the displayed colors closer to the intended colors. Interestingly, at trials where the test images were shown at standard luminance range, our method was selected 90% of the times versus the 97.5% at the high luminance range. The higher performance in the high luminance range supports our observations and measurements that the color accuracy of the HDR display is lower at high luminance levels. Nevertheless, for all the three cases we presented our method’s advantage has been found statistically significant (\( p \ll 0.01 \)).

6. CONCLUSION

We presented a systematic methodology for HDR display calibration which accounts for the variation of color reproduction errors’ dependence to luminance. Our method provides the blueprint for an HDR display calibration procedure and proposes an efficient representation for the calibration function. We demonstrate the use of this blueprint on a SIM2 HDR display, but we believe our findings and techniques will apply to other HDR displays as well. Our efficient representation of the calibration function in the form of a 3D octree forest data structure paves the way of efficient implementations both in software and hardware, which future HDR displays could leverage to make sure that content is displayed as desired.
7. REFERENCES


