# A perceptual uniformity error metric for standard and high dynamic range colour spaces

***Maryam Azimi; University of Cambridge, Cambridge, UK***
***Minjung Kim; University of Cambridge, Cambridge, UK***
***Graham D. Finlayson; University of East Anglia, Norwich, UK***
***Rafał K. Mantiuk; University of Cambridge, Cambridge, UK***

## Abstract

*In perceptually uniform colour spaces, the perceptual differences in colour pairs are approximately the same as the Euclidean distance between them. Uniformity is of great importance in applications such as gamut mapping where the perceptual difference between original and mapped colour needs to be minimised. Ideally, in a perceptually uniform colour space, the locus of constant Just Noticeable Difference (JND) around different colour samples should be the unit sphere. While several perceptually uniform colour spaces for Standard Dynamic Range (SDR) and High Dynamic Range (HDR) have been proposed, there is not a standardized uniformity metric with respect to which we might judge whether one space is more uniform than another. In this paper, we propose and develop such a uniformity metric. Importantly, our approach takes into account changes in all three directions of a colour space including luminance and this is in contradistinction to prior art that focuses mainly on the colour signal (separate from luminance). The proposed metric can be based on any perceptual colour difference metric that models JNDs.*

## Introduction

With different image processing applications emerging over the years, many colour spaces with different characteristics have been developed. Determining whether a colour space has a specific feature helps identify the appropriate colour space for a given application. Perceptual uniformity is a characteristic that is required in many applications, such as gamut mapping, lossy image compression, image enhancement, and image difference metrics, and is the focus of this work. In a perceptually uniform colour space, a change of a specific magnitude in a colour value results in a perceptual change of the same magnitude.

The perceptual colour spaces recommended by Commission Internationale de l'Éclairage (CIE) are CIE $L^*a^*b^*$ and CIE $L^*u^*v^*$ [1]. These spaces are only approximately perceptually uniform: for two different colours, the same change in colour value may give different perceptual changes. Both CIE $L^*a^*b^*$ and CIE $L^*u^*v^*$ are designed to be effective for luminance ranges of 0.1 to 100 cd/m$^2$, typical of SDR colour gamuts.

With the introduction of HDR displays, capable of reproducing a higher luminance range of 0.005 to 10,000 cd/m$^2$, the development of a perceptually uniform colour space capable of representing HDR colours is necessitated [2]. The first HDR uniform colour spaces proposed for image and video coding involved representing colour as two chromaticity coordinates CIE u'v' [3, 4] and luminance, encoded with a transfer function derived from threshold versus intensity [5] or contrast sensitivity models [6, 7]. HDR-CIELAB [8] and HDR-IPT [9] were designed to support luminance ranges up to 400 cd/m$^2$. Two other HDR-specific colour spaces are $J_za_zb_z$ [10] and $IC_TC_P$ [11] designed to be more perceptually uniform than $YC_bC_r$ [12], the most commonly used colour space for image and video compression. These colour spaces attempt to separate luminance information from chromaticity and are hence more efficient for compression purposes. Since $J_za_zb_z$ and $IC_TC_P$ colour spaces are based on the Perceptual Quantizer (PQ) [7] transfer function, they support luminance up to 10,000 cd/m$^2$.

Our goal is to quantify the perceptual uniformity of a colour space. First, we note that all commonly used colour spaces are designed to explain colour discrimination judgments; perceptual colour spaces are optimised so that the Euclidean distance in that space is approximately the same for pairs of just-discriminable colours.

One way of measuring perceived difference is to find a colour stimulus change—e.g., a perturbation in a specific colour direction—that is perceived as "**J**ust **N**oticeably **D**ifferent" to the observer. Such measurements are used to derive colour difference formulas, often denoted by $\Delta E$, which quantify the perceived difference between two arbitrary colours.

To evaluate the uniformity of colour spaces, we can simply calculate the $\Delta E$s for the pairs we have and see how well they fit the magnitude of perceived differences [10]. Indeed, this seems like a good approach and would be ideal given large enough data sets. Unfortunately, few suitable datasets are available, and even those that do exist are small, comprising only a few hundred data points [13][14][15][16]. It is a weakness of this *data-driven* approach that the perceptual uniformity of a colour space is only evaluated for the limited colour pairs available and not for all the colours present in a colour space.

In this work, our purpose is to evaluate a colour space as a whole, rather than for the colour samples with existing visual difference measurements. To this end, we set out to predict the JNDs at colours that sample the whole colour space and for all directions of colour change. In one part of our study, we will employ the perceptual colour difference metric CIEDE2000 ($\Delta E_{00}$). However, CIEDE2000 is designed only for low-dynamic-range scenes (up to 100 cd/m$^2$). Thus, we also use the HDR banding detection model [17], which allows us to infer JNDs for HDR colour spaces. This banding model is explained in more detail later.

## Related Work

Most colour difference formulas, including CIEDE2000 ($\Delta E_{00}$), are meant to quantify very small, just-noticeable colour differences, close to one JND. There have been attempts to find formulas for quantifying large, suprathreshold differences [18]. The focus of this work, however, is on the former, and hence on perceptual uniformity for JNDs.

If a Euclidean colour difference formula result fits the perceptual differences data from visual experiments, the colour

space in which the Euclidean difference formula was calculated is said to be perceptually uniform. The extent to which this Euclidean fitting does not represent the visual experimental data can be regarded as a perceptually uniformity error for that colour space. To evaluate how well the Euclidean colour difference formula results fit the visual data, the most common error metrics used are Performance Factor PF/3 [19] and the Standardized Residual Sum of Squares (STRESS) index [20]. These measures only consider small perceptual differences, which is also the scope of this work. PF/3 is a combined measure of the coefficient of variance (CV) and $\gamma$ metrics [21], and the $V_{\mathrm{AB}}$ metric [22] . One limitation of PF/3 is that it cannot indicate the statistical significance of the difference between two colour difference formulae or spaces tested. To overcome this shortcoming of PF/3 index, STRESS was proposed to allow inferences on the statistical significance of two colour difference formulae. STRESS is calculated for $N$ pairs of visual, $V_i$, and predicted (Euclidean), $E_i$, colour differences:

$$STRESS = 100\sqrt{\frac{\sum(E_i - F_1 V_i)^2}{\sum F_1^2 V_i^2}}, \tag{1}$$

where $F_1$ is chosen to best fit the predicted differences to the visual ones:

$$F_1 = \frac{\sum E_i^2}{\sum E_i V_i}. \tag{2}$$

To evaluate the perceptual uniformity of a colour space for more colour samples and directions for which visual data does not exist, one can calculate STRESS between the ellipsoids constructed from JNDs and spheres with unity radius (perfect perceptual uniformity). Two different errors can be calculated this way: local and global uniformity errors. Local uniformity means that colour discrimination ellipsoids should be spheres. Global uniformity means that the area of all spheres should be equal. Fig. 1 shows an example illustrating the ideas of local and global uniformity errors. Local uniformity error is zero for Fig. 1(a) and (b) as both of the shapes are spheres although with different radius. However, global uniformity error is zero for Fig. 1(a) whilst it results in an error for Fig. 1(b). As for Fig. 1(c), both local and global uniformity errors are larger than zero.

For the local uniformity test, STRESS can be computed between ratios of the principal axes of ellipsoids and a unity vector (ratio of the principal axes of a sphere). To test global uniformity, STRESS can be computed between the areas of the ellipsoids and a constant value represented by the mean area. For a colour space to be perfectly perceptually uniform both local and global uniformity should be zero.

This approach of using STRESS to evaluate a colour space's uniformity based on local and global uniformity, was performed in [10]. However, in that study, changes in luminance directions were not considered and hence ellipsoids and spheres were changed to ellipses and circles, respectively. Of course, a problem with the local and global approach is precisely that two separate measures are produced. How we combine the individual global and local measures into a single figure of merit is not clear.

## Perceptual colour difference formulae

Colour difference formulae are designed to detect the JND, the minimum noticeable difference between a reference colour and a second colour that is a small offset away from the reference. With the development of CIE L*a*b* as a perceptually uniform colour space, $\Delta E^*_{\mathrm{ab}}$ could be calculated as an Euclidean distance between two points. If Euclidean distance calculated in CIE L*a*b* perfectly correlated with perceptual colour differences, then equal colour difference data plotted in CIE L*a*b* coordinates would lie on a sphere. However, as reported in [23], colour changes with equal tolerance are represented in CIE L*a*b* by ellipsoids of variable size instead of same-size spheres. It follows then that Euclidean distance in CIE L*a*b*, denoted $\Delta E^*_{\mathrm{ab}}$ is not a perfectly perceptual colour difference metric.

To understand whether possible discrepancies of the colour difference dataset caused the JNDs to form ellipsoids rather than spheres in CIE L*a*b* colour space, new datasets with well-controlled conditions compared to earlier datasets are evaluated in [24]. It is reported in [24] that even with the well-controlled new visual dataset, when the lightness $(L^*)$ component is fixed, the colour ($a^*$ and $b^*$) JND ellipses for grey colours are smaller than those of the other colours. Furthermore, the JND ellipses for blue colours are differently oriented compared to those of the other colours. Consequently, due to these differences in the size and orientation of the $a^*$ and $b^*$ JND ellipses for different colours, CIE L*a*b* is not a perceptually uniform colour space.

To address the non-uniformity of CIE L*a*b*, CIEDE2000 metric $(\Delta E_{00})$ is developed as a non-Euclidean colour difference formula with new weighting functions for hue and lightness components [24]. A key subtlety here is that $\Delta E_{00}$ is a formula and not a colour space (we do not have a coordinate transform with respect to which we calculate the Euclidean distance). More importantly, $\Delta E_{00}$ provides a much better fit of the visual experimental data compared both to distances calculated in CIE L*a*b* and compared with preceding colour difference formulae such as $\Delta E_{\mathrm{CMC}}$ and $\Delta E^*_{94}$.

An important aspect of most colour spaces and colour difference formulae is that they are designed to work with specific viewing conditions. For example, the $\Delta E_{00}$ formula was derived for specific reference viewing conditions that included the white point set to D65 with a maximum scene luminance of 100 cd/m$^2$. To employ $\Delta E_{00}$ for different conditions, such as extending it to HDR luminance levels, the $\Delta E_{00}$ weighting factors need to be newly derived. Indeed, merely changing the white point to higher luminance values is not sufficient. Therefore, $\Delta E_{00}$ might not generalise as a perceptual colour difference metric for HDR colours.

With HDR displays a wider range of colours at higher and lower luminance values are introduced to the gamut reproduced by new displays. In [25], an HDR colour difference dataset is gathered and JND values are marked for the tested colours. Colour difference metric of $\Delta IC_TC_P$ is proposed in [26] to specify these JNDs. The $\Delta IC_TC_P$ is the Euclidean distance in $IC_TC_P$, where the $C_p$ component is halved before calculating the distance. It is reported in [25] that the proposed colour difference

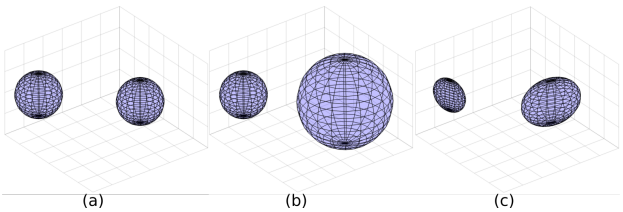


***Figure 1.*** *Colour discrimination ellipsoids for an imaginary colour space to show local and global uniformity errors: (a) with zero local and global uniformity errors, (b) with zero local uniformity error and non-zero global uniformity error, and (c) non-zero local and global uniformity errors.*

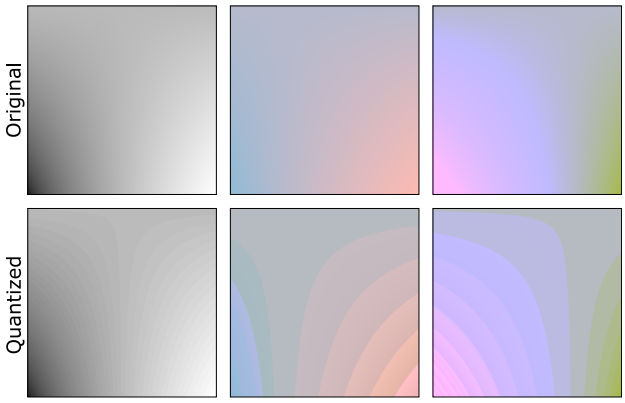


***Figure 2.*** *Smooth gradients (top) in greyscale and colour reveal banding artifacts (bottom) when represented with insufficient bit-depth.*

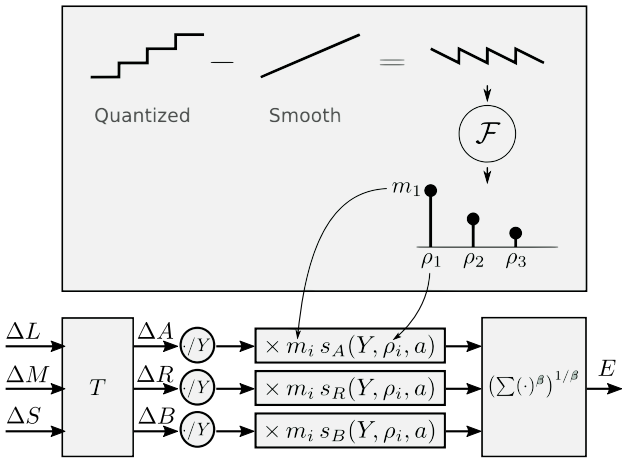


***Figure 3.*** *Banding detection model. The difference between the quantized and smooth (unquantized) signals is transformed into the Fourier domain to find the spatial frequencies ($\rho_i$) and amplitudes ($m_i$) of the banding artifacts (top box). Those are used to find a detection threshold using an energy model (bottom blocks) operating on a colour-opponent signals ($\Delta A$, $\Delta R$, $\Delta B$), that are normalized by the luminance $Y$. $\Delta R$ refers to L-M difference (red-green) and $\Delta B$ refers to L-S difference (yellow-violet). $T$ is a $3\times3$ matrix transforming cone responses into a colour opponent space (DKL). The signal for each colour and frequency component is modulated by the spatio-chromatic CSF ($s_{\{A,R,B\}}(\cdot)$) from [27], which is a function of luminance $Y$, spatial frequency $\rho$ and stimulus size $a$.*

metric performs better compared to $\Delta E_{00}$ at the tested luminance levels in predicting the JND ellipses in chromaticity (i.e., when luminance differences are not considered).

Another HDR colour difference metric is $\Delta E_z$ which is the Euclidean distance between two colours represented by $J_z a_z b_z$, a colour system designed to be perceptually uniform for HDR wide colour and high luminance range. The performance of $\Delta E_z$ is evaluated in [26] compared to that of $\Delta IC_TC_P$ in predicting the JND ellipses in chromaticity, not considering luminance changes. It is reported in [26] that $\Delta E_z$ outperforms $\Delta IC_TC_P$ for the tested colors, all of which were below 100 cd/m$^2$ in luminance.

### *HDR colour banding detection model*

The majority of lossy image and video compression methods require representing colours as integer values of a fixed bit-depth. Such a conversion to integer numbers may result in colour banding artifacts, also known as contouring. Examples of such artifacts are shown in Fig. 2. This problem begs the question of how we detect colour banding in images. Colour banding detection models can be used, in general, to evaluate the visibility of banding artefacts. Here we use a colour banding model to test whether a colour space is suitable for efficient image and video coding using lossy compression. In [17] it is shown that the visibility of chromatic banding artifacts in SDR content can be predicted with a model that analyzes the banding signal in the Fourier domain. Here, we extend our model in [17] to predict banding in HDR content. The model is based on a new spatio-chromatic CSF for a large range of absolute luminance from [27]. Since our work is based on [17], we will now describe this prior art in more detail.

The processing diagram of the banding model can be found in Fig. 3. The model takes as input the average colour of the gradient (as LMS cone responses) and the direction in which the colours change within the gradient ($\Delta L$, $\Delta M$, $\Delta S$). It predicts the height of the gradient step that is just noticeable (75% probability of detection in an 2-alternative-forced-choice (2AFC) experiment). The model predicts banding for the worst-case stimulus; a smooth gradient in greyscale or colour, such as the one shown in Fig. 2. The quantization artifacts in such a gradient form a *staircase* pattern. The difference between the staircase pattern and a smooth gradient forms a *saw-tooth* pattern, which has an analytical form of its Fourier transform. This allows to decompose the difference signal into multiple frequency components and into three colour opponent channels, weight each by the spatio-chromatic CSF [27] and then accumulate the energy across all bands and colour channels. If such cumulative energy exceeds a pre-determined threshold, the banding artifacts are predicted to be visible. More details on the banding model can be found in [17].

The model in [17] is calibrated and tested on a dataset that includes gradients in luminance and chrominance, displayed on a DayDream Virtual Reality (VR) headset. It relied on a simplified empirical spatio-chromatic CSF. Here, we extend the model so that it can predict thresholds in the Rec.2020 colour gamut of HDR colour spaces. This is achieved by replacing the original empirical CSF with the spatio-chromatic CSF from [27]. The model was tested on the data from both the original publication [17] and the set of data measured on an HDR display. The details of the model are presented in [17].

## Proposed perceptual uniformity metric for colour spaces

Fig. 4 shows the JND ellipses using $\Delta E_{00}$ and the HDR colour banding detection model for HDR colour spaces of linear Rec.2020 RGB, PQ Rec.2020 RGB, PQ Rec.2020 $YC_bC_r$, $IC_TC_P$ and $J_z a_z b_z$. Fig. 5 shows the same ellipses for SDR colour space of linear Rec.709 RGB, gamma Rec.709 RGB, gamma Rec.709 $YC_bC_r$, CIE $L^*u^*v^*$ and CIE $L^*a^*b^*$. Value of 2.2 was used for gamma encoding. Note that the left side of Fig. 4 and Fig. 5 show the second and third dimensions of each space while the first dimension is constant. The right side of both figures visualizes the ellipses when the third dimension is constant. It can be seen from Fig. 4 that in PQ Rec.2020 RGB and PQ Rec.2020 $YC_bC_r$, (1) most of the ellipses are not circular and (2) the ellipses vary drastically in their size. That suggests that PQ Rec.2020 RGB and PQ Rec.2020 $YC_bC_r$ could be less perceptually uniform than $IC_TC_P$ and $J_z a_z b_z$. However, those ellipses show only a small sample of the colour gamut and it is unclear how much better $IC_TC_P$ and $J_z a_z b_z$ are performing compared to PQ Rec.2020 RGB and PQ Rec.2020 $YC_bC_r$ in

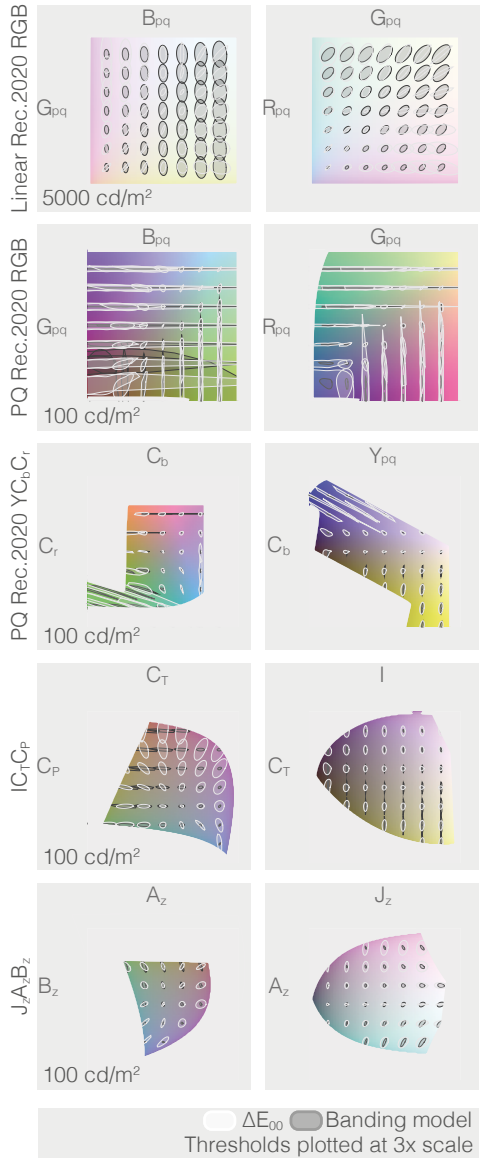


***Figure 4.*** *Perceptual uniformity thresholds in terms of $\Delta E_{00}$ and the banding model in [17] for linear Rec.2020 RGB, PQ Rec.2020 RGB, PQ Rec.2020 $YC_bC_r$, $IC_TC_P$ and $J_za_zb_z$.*

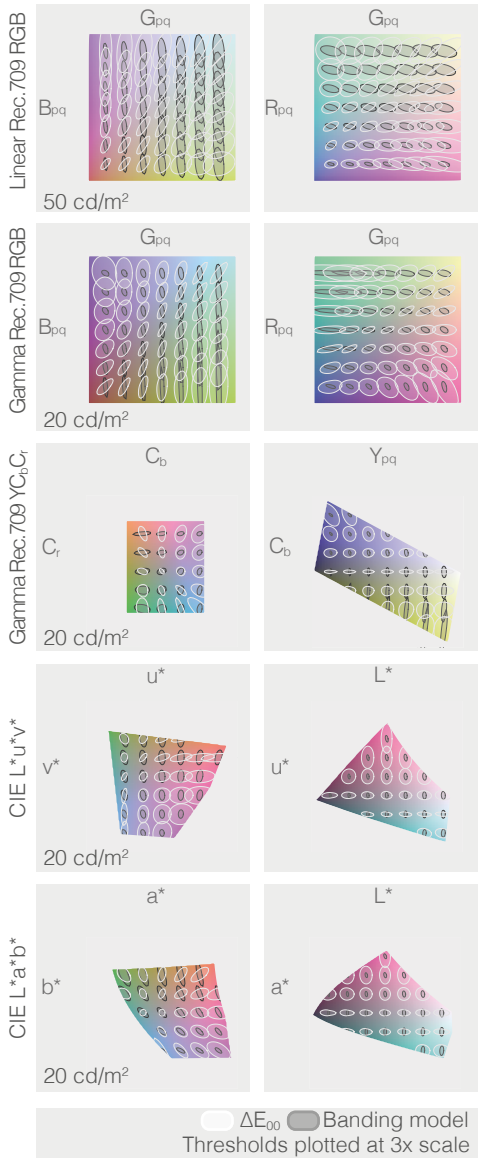


***Figure 5.*** *Perceptual uniformity thresholds in terms of $\Delta E_{00}$ and the banding model in [17] for linear Rec.709 RGB, Gamma Rec.709 RGB, Gamma Rec.709 $YC_bC_r$, CIE $L^*u^*v^*$ and CIE $L^*a^*b^*$.*

terms of perceptual uniformity. Same observation holds for PQ Rec.709 RGB and PQ Rec.709 $YC_bC_r$ compared to CIE $L^*a^*b^*$ and CIE $L^*u^*v^*$ in Fig. 5.

Here, we propose a single error metric that can quantify the perceptual uniformity errors of a colour space. Note that while Fig. 4 and Fig. 5 show JNDs in 2D for better visualization, our metric estimates perceptual non-uniformity in all three dimensions.

Our error measure estimates the error when an Euclidean distance is used in a given colour space to predict a JND. Any existing colour difference formulae can be employed to find the distance corresponding to 1 JND from a sample point $p_i$ in the direction defined by spherical coordinates $\phi$ and $\theta$ in the colour space. We denote such a 1-JND distance as $r(p_i, \theta, \phi)$. Then, we can compute the expected value of errors as:

$$\varepsilon = \frac{1}{N}\sum_{i=1}^{N} \mathbb{E}_{\theta,\phi}\left[\left|\log_2 \frac{r(p_i,\theta,\phi)}{r_0}\right|\right], \tag{3}$$

where $\mathbb{E}_{\theta,\phi}[\cdot]$ is the expected value over all directions $\phi$ and $\theta$ in the colour space. $r_0$ is the distance that minimizes the expected error measure $\varepsilon$, which can be found by computing a geometric mean over all directions and sample points $p_i$:

$$\log_2(r_0) = \frac{1}{N}\sum_{i=1}^{N} \log_2\left(r(p_i,\theta,\phi)\right) \tag{4}$$

We express the error as a logarithmic ratio to capture how much the ellipsoids need to be deformed to make the colour space uniform. The error of 1 unit means that, on average, the JND estimated as an Euclidean distance is twice as large or twice as small as the distance predicted by a colour difference metric. A perfectly uniform colour space, both locally and globally (according to a given colour difference metric), will result in $\varepsilon = 0$. We provide the Matlab code of our uniformity metric[1].

[1] Matlab code for the uniformity metric:
https://github.com/gfxdisp/colourspace-uniformity-metric

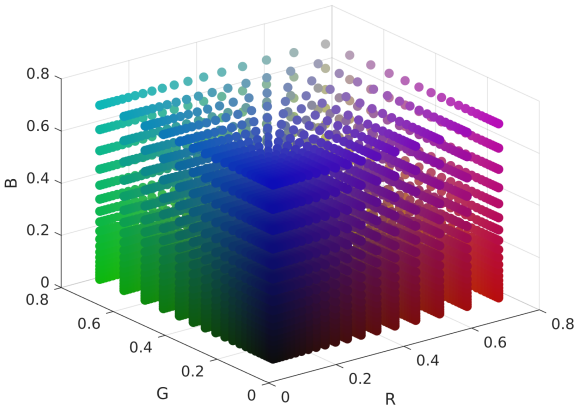


***Figure 6.*** *Colour samples $p_i$ in RGB grid with $20 \times 20 \times 20$ samples.*

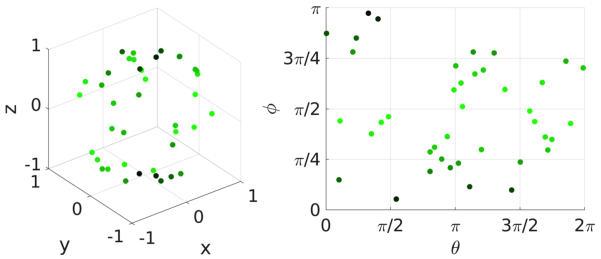


***Figure 7.*** *Uniform spherical samples used to generate $(\theta, \phi)$.*

## Experimental Setup

Using the metric proposed in the previous section, we evaluated the Gamma Rec.709 RGB, Gamma Rec.709 $YC_bC_r$, CIE $L^*a^*b^*$, and CIE $L^*u^*v^*$ colour spaces in terms of their perceptual uniformity for SDR colours. For HDR colours, we evaluated PQ Rec.2020 RGB, PQ Rec.2020 $YC_bC_r$, $IC_TC_P$, and $J_za_zb_z$. We also added linear RGB colour space to our test sets as a colour space that is in essence non-perceptual and hence perceptually non-uniform to verify our metric's ability to pick up such extreme case in a colour space. Note that we scaled all three components of each colour space so that they are all of the range [0, 1] such that the relative ratios between components values are maintained.

We sample $50 \times 50 \times 50$ colours $p_i$ that form a grid in a linear RGB colour space. We use ITU Rec.709 colour primaries with a peak luminance of 100 cd/m$^2$ and a black level of 0.1 cd/m$^2$ for SDR colour spaces, and ITU Rec.2020 colour primaries with a peak luminance of 10,000 cd/m$^2$ and a black level of 0.005 cd/m$^2$ for HDR colour spaces. The points of the grid are distributed logarithmically along R, G and B axes to account for Weber's law. The start and end points are offset by 10% of the R, G or B value from the boundary of the gamut. This is because most colour difference formula cannot provide robust predictions for colours close to the gamut boundary. Fig. 6 shows how our colour samples are distributed in the RGB grid.

For directions $(\theta, \phi)$, samples are generated by uniform spherical sampling in RGB through generating a random point on the cylinder [- 1, 1]x[0, $2\pi$] and then finding its inverse axial projection on the unit sphere, as explained in [28]. Fig. 7 shows 40 uniform spherical direction samples.

Overall, 40 (directions)$\times$125,000 (colour samples) resulting in 5,000,000 $r(p_i, \theta, \phi)$ values are used to generate our results. The values of 125,000 samples for $N$ (colour samples) and 40 for direction samples are selected as increasing these values did not substantially change the metrics result. The JND distances $r(p_i, \theta, \phi)$ are initially computed in the linear RGB and then converted to the target colour spaces. This is because some of the colour difference metrics offer closed-form estimation of JND distances in a linear colour space but require slow binary search in non-linear spaces.

## Results and discussions

Table 1 reports the perceptual uniformity errors for the tested HDR and SDR colour spaces in terms of the proposed metric. We calculated JND distances $r(p_i, \theta, \phi)$ based on $\Delta E_{00}$ and the HDR colour banding detection model as explained earlier. Note that $\Delta E_{00}$ = 1 was used as the JND value to form JND distances $r(p_i, \theta, \phi)$.

As expected linear Rec.2020 and Rec.709 RGB colour spaces yield the largest perceptual uniformity error in terms of the proposed metric compared to the other tested HDR and SDR colour spaces, which include non-linear coding. This is because linear RGB values are strongly perceptually non-uniform as they do not account for Weber's law (luminance masking). Such non-uniformity of the linear RGB shows how ill-suited it is as a basis to calculate objective image quality metrics on (e.g. PSNR or SSIM [29] ). For an objective metric results to be correlated with perceptual measures, it is crucial that the metric is applied on a perceptually uniform colour space.

PQ Rec.2020 RGB and $YC_bC_r$ yield smaller errors in terms of perceptual uniformity compared to linear RGB as they apply PQ on all three components. There is no clear advantage for $YC_bC_r$ in terms of perceptual uniformity over PQ Rec.2020 RGB; the error for $YC_bC_r$ is even larger compared to all cases except when $\Delta E_{00}$ metric is used for the HDR colour space. This lack of gain in uniformity of $YC_bC_r$, as compared to RGB, can be due to the Non-constant Luminance (NCL) derivation of $Y$ component in $YC_bC_r$ [30]. However, since $YC_bC_r$ isolates the luminance from chromaticity information, it is preferred for image and video compression applications as it enables subsampling the chromaticity components as well as taking advantage of the reduced redundancy between the components. $IC_TC_P$ also isolates luminance from chromaticity and is also more perceptually uniform than $YC_bC_r$ for HDR colours.

$J_za_zb_z$ yields higher perceptual uniformity error compared to $IC_TC_P$. This observation is in conflict with the results in [10] where reportedly $J_za_zb_z$ is more perceptually uniform that $IC_TC_P$ in terms of global and local uniformity measures used in [10]. One reason could be that in [10], uniformity errors are only evaluated for chromaticity channels and the luminance channel changes are not considered. While $J_za_zb_z$ performs well in maintaining chromaticity uniformity, by adding luminance dimension, the performance changes. Another explanation for our result (in conflict with the previous study) may be because the colour samples used in [10] to calculate uniformity errors are, perforce, different from the more comprehensive set of samples we use. In [10], errors are calculated for a limited number of colour samples present in COMBVD (125 colours) [10] and MacAdam (25 colours) [13] datasets while in this study we logarithmically sample the RGB cube with $50 \times 50 \times 50$ = 125,000 colours. Furthermore, the maximum luminance of colours in datasets used in [10] are 100 cd/m$^2$ while our results are calculated for colours up to 10,000 cd/m$^2$ luminance.

In the case of SDR colour spaces, when $\Delta E_{00}$ is used for constructing JND ellipsoids, CIE $L^*a^*b^*$ colour space is found to be the most perceptually uniform. This may be expected

**Perceptual uniformity errors in HDR (Rec.2020) and SDR (Rec.709) colour spaces based on the proposed error metric with uniformity ellipsoids based on $\Delta E_{00}$ and HDR banding detection model, as well as the local and global uniformity errors using STRESS with uniformity ellipsoids based on HDR banding detection model.**

| Metric | HDR Colour spaces – BT.2020 | | | | | SDR Colour spaces – BT.709 | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Linear RGB | PQ RGB | PQ $YC_bC_r$ | $IC_TC_P$ | $J_za_zb_z$ | Linear RGB | Gamma RGB | Gamma $YC_bC_r$ | CIE $L^*u^*v^*$ | CIE $L^*a^*b^*$ | IPT | $J_za_zb_z$ |
| Proposed $\varepsilon$ ($\Delta E_{00}$) | 1.867 | 0.782 | 0.746 | **0.518** | 0.662 | 0.826 | 0.591 | 0.604 | 0.513 | **0.370** | 0.476 | 0.375 |
| Proposed $\varepsilon$ (HDR Banding Detection Model) | 2.282 | 0.968 | 1.132 | **0.853** | 0.935 | 0.901 | 0.491 | 0.495 | **0.388** | 0.410 | 0.409 | 0.456 |
| STRESS local (HDR Banding Detection Model) | 83.03 | 51.49 | 60.71 | **26.54** | 31.89 | 19.08 | 50.77 | 38.96 | 18.53 | 17.97 | 17.13 | **15.89** |
| STRESS global (HDR Banding Detection Model) | 91.44 | 87.40 | 86.15 | **36.41** | 76.83 | 85.88 | 87.95 | 86.82 | 59.67 | 58.95 | 58.51 | **53.39** |
| KL Divergence (HDR Banding Detection Model) | 25.50 | **17.15** | 17.76 | 17.95 | 20.20 | 17.04 | **14.19** | 15.06 | 16.82 | 16.25 | 16.19 | 16.02 |

as $\Delta E_{00}$ is calculated and designed based on this colour space. However, when the HDR colour banding detection model is used, CIE $L^*u^*v^*$ yields the smallest perceptual uniformity error. Similar to HDR case, there is no clear advantage in terms of perceptual uniformity in using $YC_bC_r$ over gamma Rec.709 RGB.

Another observation from comparing the perceptual uniformity errors is that the banding detection model and $\Delta E_{00}$ yield similar results for SDR colour spaces, but very different results for HDR colour spaces (with the exception of CIE $L^*a^*b^*$, which results in different error for both metrics). However, in the case of HDR colour spaces, the HDR colour banding detection model results in higher perceptual uniformity errors than $\Delta E_{00}$. This could be due to the fact that $\Delta E_{00}$ is calibrated for the limited brightness range of SDR while the banding model is devised from the measurements of up to 10,000 cd/m$^2$. This indicates that $\Delta E_{00}$ might not generalize to the HDR colour spaces.

While the results in Table 1 report the average perceptual uniformity errors, we show the distributions of these errors for HDR colour spaces in Fig. 8 and for SDR colour spaces in Fig. 9, using a colour bar. These errors are calculated based on the expected error for each colour, sampled in the given colour space. Note that the samples used to generate these figures are different from the ones used to calculate the results in Table 1, which were sampled in the log-RGB space. The JNDs were constructed from the HDR colour banding detection model. For visualization purpose, the errors are shown in 2D for the 2nd and 3rd dimension of each colour space while their first component is kept at a constant value, reported on top of each slice. As can be observed from Figs. 8 and 9, the perceptual uniformity errors are larger for colours with higher chromaticity values (more towards the border of the colour gamuts). For colours at the center of the gamut (colours closer to white), perceptual uniformity errors are lower. This implies that while perceptual transfer functions such as PQ and gamma encoding can play an important role in making a colour space perceptually uniform, it mainly affects only the uniformity of colours that are around white. PQ was originally designed for greyscale colours and based on a luminance-only CSF. While applying PQ on chromaticity channels is convenient, a chromaticity perceptual transfer function that is derived specifically based on a chromaticity CSF could further benefit the colour spaces in terms of perceptual uniformity.

### *Prior-art uniformity test*

We also evaluated the perceptual uniformity of the tested colours spaces in terms of their local and global perceptual uniformity as explained in "Related Work". We follow the workflow in [10] and extend it to 3-dimensional colour spaces, as perceptual uniformity was evaluated in [10] only in terms of chromaticity. We first fit an ellipsoid to the JNDs obtained from the HDR colour banding detection model for each of the $N$ colour samples $p_i$. Let $A_i$, $B_i$, and $C_i$ represent the principal axes of each ellipsoid. Recall that local uniformity evaluates how close each ellipsoid is to a sphere. To test the local uniformity, STRESS was computed using Eq.( 1) between

$$\Delta E = \left[ \frac{\frac{A_1}{B_1}+\frac{A_1}{C_1}+\frac{B_1}{C_1}}{3} \quad \frac{\frac{A_2}{B_2}+\frac{A_2}{C_2}+\frac{B_2}{C_2}}{3} \quad \dots \quad \frac{\frac{A_N}{B_N}+\frac{A_N}{C_N}+\frac{B_N}{C_N}}{3} \right], \quad (5)$$

which is a vector containing the ratios of the $N$ ellipsoids' principal axes and

$$\Delta V = \begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix}, \quad (6)$$

which is a vector containing the ratios of $N$ spheres principal axes resulting in a unity vector with $N$ elements.

Recall that global uniformity evaluates how similar the sizes of all the ellipsoids are to each other. To test global uniformity,

STRESS was computed between

$$\Delta E = \Bigg[ 4\pi \left( \frac{(A_1B_1)^{1.6} + (A_1C_1)^{1.6} + (B_1C_1)^{1.6}}{3} \right)^{1/1.6} \quad 4\pi \left( \frac{(A_2B_2)^{1.6} + (A_2C_2)^{1.6} + (B_2C_2)^{1.6}}{3} \right)^{1/1.6} \quad \dots \quad 4\pi \left( \frac{(A_NB_N)^{1.6} + (A_NC_N)^{1.6} + (B_NC_N)^{1.6}}{3} \right)^{1/1.6} \Bigg], \tag{7}$$

a vector containing surface areas of the $N$ ellipsoids and

$$\Delta V = \left[ \overline{\Delta E} \quad \overline{\Delta E} \quad \dots \quad \overline{\Delta E} \right], \tag{8}$$

a vector containing $N$ constant values representing the mean surface area (all ellipsoids equal in size).

The local and global uniformity errors are reported on the right side of Table 1. In the majority of cases the global uniformity metrics yield similar results as the proposed metric. However, the local uniformity metrics are very different from the proposed metrics results. A colour space that is perceptually uniform in terms of global uniformity can be otherwise in terms of the local metric. For instance, RGB linear results in very low local uniformity error while having the highest global uniformity error. This behaviour is well visible in Figs. 4 and 5 where linear spaces have the least elongated ellipses (low local error) while these ellipses differ significantly in size (high global error). One advantage of the proposed metric is that it encompasses both of the local and global uniformity measures in one single value that is easier to interpret.

### *Validation*

To further validate the proposed metric, we refer to experimental evaluations performed in [31], where it was reported based on subjective evaluations that CIE L*u*v* colour space is less perceptually uniform compared to PQ Rec.2020 $YC_bC_r$. Specifically, the minimum number of bits required to represent CIE L*u*v* and PQ Rec.2020 $YC_bC_r$ without banding artefacts were identified at luminance levels from 0.05 to 50 cd/m$^2$. It was shown that for PQ Rec.2020 $YC_bC_r$, the required number of bits for the signal to be represented without visible banding artefacts does not change when luminance is increased. In contrast, CIE L*u*v* requires fewer bits in dark areas compared to bright. Hence, it was concluded that PQ Rec.2020 $YC_bC_r$ is more perceptually uniform. Please note that PQ was used as the transfer function of the L component in CIE Lu'v', therefore we denote that space as $Y_{PQ}$u'v'. We apply our proposed perceptual uniformity error metric on $Y_{PQ}$u'v' and PQ Rec.2020 $YC_bC_r$ to see whether the metric yields the same results as in [31]. For colour samples, we take $20 \times 20 \times 20$ colours $p_i$ that form a grid in a linear RGB colour space. We use ITU Rec.2020 colour primaries with a peak luminance of 50 cd/m$^2$ and a black level of 0.05 cd/m$^2$ to be consistent with the experiments performed in [31] . The points of the grid are distributed logarithmically along R, G and B axes to account for Weber's law. The start and end points are offset by 10% of the R, G or B value from the boundary of the gamut. As for directions, we generate 40 uniform spherical samples. The results are reported in Table 2 for both the proposed metric and the local and global uniformity metrics based on STRESS. We only report the results in terms of the HDR colour banding detection model as the minimum luminance of the experiments (0.05 cd/m$^2$) is in the range of HDR colours.

**Perceptual uniformity error in PQ Rec.2020 $YC_bC_r$ and CIEL'u'v' for luminance ranges of 0.05 to 50 cd/m$^2$ based on the proposed error metric. The local and global uniformity errors using STRESS with uniformity ellipsoids based HDR banding detection model are also reported.**

| Metric | Colour spaces | |
|---|---|---|
| | PQ $YC_bC_r$ | $Y_{PQ}$u'v' |
| Proposed $\varepsilon$ (HDR Banding Detection Model) | **0.8151** | 0.8485 |
| STRESS local (HDR Banding Detection Model) | **69.40** | 86.09 |
| STRESS global (HDR Banding Detection Model) | **70.02** | 78.38 |

Also, in [31] the experiments were performed to detect banding artefacts and hence it is more meaningful to use an objective banding detection model to match the visual results in [31]. It can be noted that the proposed metric evaluated PQ Rec.2020 $YC_bC_r$ to be more perceptually uniform than $Y_{PQ}$u'v', which is consistent with the visual experiments results in [31].

## Conclusions

In this paper, we proposed a perceptual uniformity error metric which can be used to compare colour spaces based on their perceptual uniformity. As opposed to prior work, our metric evaluates the local and global uniformity of a colour space with one single value, and hence offers simpler and more conclusive interpretation. The metric relies on a perceptual colour difference metric that models JNDs, such as $\Delta E_{00}$ or a colour banding detection model. This enables the metric to uniformly sample the tested colour space and avoid bias due to a dataset, which could cover only a portion of the space. This is especially important for HDR colour spaces, for which the available observer data is very limited. Importantly, our approach considers colour change in all three directions of a colour space including luminance. Our 3-dimensional approach is in contradistiction to the prior art that focuses mostly the uniformity of chromaticity. Our results showed that the metric can robustly rate colour spaces based on their uniformity, consistent with psychophysical experiments.

## References


[1] Commission Internationale de l'Éclairage (CIE). Colorimetry, cie publication no. 15. 2004.

[2] Rafał K. Mantiuk, Karol Myszkowski, and Hans-Peter Seidel. *High Dynamic Range Imaging*, pages 1–42. (2015).

[3] Gregory Ward-Larson. LogLuv Encoding for Full-Gamut, High-Dynamic Range Images. *Journal of Graphics Tools*, 3(1):15–31, Jan (1998).

[4] Greg Ward and Maryann Simmons. JPEG-HDR: A Backwards-Compatible, High Dynamic Range Extension to JPEG. In *Color and Imaging Conference Conference*, pages 283–290(8), New York, New York, USA, (2005). ACM Press.

[5] Rafał Mantiuk, Grzegorz Krawczyk, Karol Myszkowski, and Hans-Peter Seidel. Perception-motivated high dynamic range video encoding. *ACM Transactions on Graphics*, 23(3):733, Aug (2004).

[6] Rafał Mantiuk, Karol Myszkowski, and Hans-Peter Seidel. Lossy Compression of High Dynamic Range Images and Video. In *Human Vision and Electronic Imaging*, page 60570V, (2006).

[7] Scott Miller, Mahdi Nezamabadi, and Scott Daly. Perceptual signal coding for more efficient usage of bit codes. In *The 2012 Annual Technical Conference Exhibition*, pages 1–9, (2012).

[8] Mark D. Fairchild and David R. Wyble. HDR-CIELAB and HDR-IPT: Simple models for describing the color of high-dynamic-range and wide-color-gamut images. *Final Program and Proceedings - IS and T/SID Color Imaging Conference*, pages 322–326, 01 (2010).

[9] Mark D. Fairchild and Ping-Hsu Chen. Brightness, lightness, and specifying color in high-dynamic-range scenes and images. In Susan P. Farnand and Frans Gaykema, editors, *Image Quality and System Performance VIII*, volume 7867, pages 233 – 246. International Society for Optics and Photonics, SPIE, (2011).

[10] Muhammad Safdar, Guihua Cui, Youn Jin Kim, and Ming Ronnier Luo. Perceptually uniform color space for image signals including high dynamic range and wide gamut. *Optics express*, 25(13):15131–15151, (2017).

[11] Taoran Lu, Fangjun Pu, Peng Yin, Tao Chen, Walt Husak, Jaclyn Pytlarz, Robin Atkins, Jan Fr-hlich, and Guan-Ming Su. Itp colour space and its compression performance for high dynamic range and wide colour gamut video distribution. *ZTE Communications*, 14(1):32–38, (2019).

[12] BT Series. Worldwide unified colorimetry and related characteristics of future television and imaging systems. 1998.

[13] David L MacAdam. Visual sensitivities to color differences in daylight. *JOSA*, 32(5):247–274, (1942).

[14] Roy S Berns, David H Alman, Lisa Reniff, Gregory D Snyder, and Mitchell R Balonon-Rosen. Visual determination of suprathreshold color-difference tolerances using probit analysis. *Color Research & Application*, 16(5):297–316, (1991).

[15] Klaus Witt. Geometric relations between scales of small colour differences. *Color Research & Application*, 24(2):78–92, (1999).

[16] Fritz Ebner and Mark D Fairchild. Finding constant hue surfaces in color space. In *Color Imaging: Device-Independent Color, Color Hardcopy, and Graphic Arts III*, volume 3300, pages 107–117. International Society for Optics and Photonics, (1998).

[17] Gyorgy Denes, George Ash, Huameng Fang, and Rafał K Mantiuk. A visual model for predicting chromatic banding artifacts. In *Human Vision and Electronic Imaging*, volume 31, pages 1–8. Society for Imaging Science and Technology, 2019.

[18] Klaus Richter. Cube-root color spaces and chromatic adaptation. *Color Research & Application*, 5(1):25–43, (1980).

[19] Shing-Sheng Guan and M Ronnier Luo. Investigation of parametric effects using small colour differences. *Color Research & Application*, 24(5):331–343, (1999).

[20] Pedro A. García, Rafael Huertas, Manuel Melgosa, and Guihua Cui. Measurement of the relationship between perceived and computed color differences. *J. Opt. Soc. Am. A*, 24(7):1823–1829, Jul (2007).

[21] C Alder, KP Chaing, TF Chong, E Coates, AA Khalili, and B Rigg. Uniform chromaticity scales—new experimental data. *Journal of the Society of Dyers and Colourists*, 98(1):14–20, (1982).

[22] W. Schultze. The usefulness of colour-difference formulae for fixing colour tolerances. In *Proceedings AIC/Holland (Soesterberg, 1972)*, pages 254–265, (1972).

[23] Roderick McDonald. Industrial pass/fail colour matching part 1-preparation of visual colour-matching data. *Journal of the Society of Dyers and Colourists*, 96:372 – 376, 10 (2008).

[24] M Ronnier Luo, Guihua Cui, and Bryan Rigg. The development of the CIE 2000 colour-difference formula: CIEDE2000. *Color Research & Application*, 26(5):340–350, 2001.

[25] Jaclyn Pytlarz and Elizabeth Pieri. How close is close enough? specifying colour tolerances for HDR and WCG displays. *IET journals*, (2017).

[26] Elizabeth Pieri and Jaclyn Pytlarz. Hitting the mark - a new color difference metric for HDR and WCG imagery. In *SMPTE 2017 Annual Technical Conference and Exhibition*, pages 1–13, (2017).

[27] Sophie Wuerger, Maliha Ashraf, Minjung Kim, Jasna Martinovic, María Pérez-Ortiz, and Rafał K. Mantiuk. Spatio-chromatic contrast sensitivity under mesopic and photopic light levels. *Journal of Vision*, 20(4):23, Apr (2020).

[28] Min-Zhi Shao and Norman Badler. Spherical sampling by archimedes' theorem. *Technical Reports (CIS)*, page 184, (1996).

[29] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, (2004).

[30] BT Series. Parameter values for ultra-high definition television systems for production and international programme exchange. In *Proc. ITU-T, BT.2020*, pages 1–7, 2012.

[31] Ronan Boitard, Rafał K. Mantiuk, and Tania Pouli. Evaluation of Color Encodings for High Dynamic Range Pixels. In *Proc. SPIE 9394, Human Vision and Electronic Imaging XX*, San Francisco, (2015).

## Author Biography

*Maryam Azimi is an Aesthetics Algorithms Scientist at Apple. Prior to that, she was a postdoctoral research fellow of the Natural Sciences and Engineering Research Council of Canada with the Department of Computer Science and Technology, University of Cambridge. She received her MASc and PhD degrees in electrical and computer engineering from the University of British Columbia, Vancouver, Canada, in 2014 and 2019, respectively. She is an active member of the Standard Council of Canada, MPEG and the IEEE Signal Processing Society.*

*Minjung Kim is a postdoctoral research associate at the Department of Computer Science and Technology, University of Cambridge. Her PhD is from New York University and York University, Canada. Her research interests include the perception of colour, light, and shape, and the computational modeling of visual perception.* `https://www.minjung.ca`

*Graham Finlayson is a professor of computing sciences at the University of East Anglia where he is also director of their Colour & Imaging lab. Graham's research spans colour image processing, physics-based computer vision and visual perception. As well as developing the underlying theory and algorithms, Professor Finlayson has a long track record of commercialising his technology including in his successful spinout companies Imsense and Spectral Edge Ltd. He is a fellow of the Institute of Engineering Technology, the Society of Imaging Science and Technology and the Royal Photographic Society.*

*Rafał K. Mantiuk is a Professor of Graphics and Displays at the Department of Computer Science and Technology, University of Cambridge (UK). He received a Ph.D. from the Max-Planck Institute for Computer Science (Germany). His recent interests focus on computational displays, rendering, and imaging algorithms that adapt to human visual performance and deliver the best image quality given limited resources, such as computation time or bandwidth. He contributed to early work on high-dynamic-range imaging, including quality metrics (HDR-VDP), video compression, and tone mapping. More recently, he led an ERC-funded project on a capture-and-display system that passed the visual Turing test: 3D objects were reproduced with sufficient fidelity to be indistinguishable from their real counterparts. Further details: http://www.cl.cam.ac.uk/ rkm38/.*

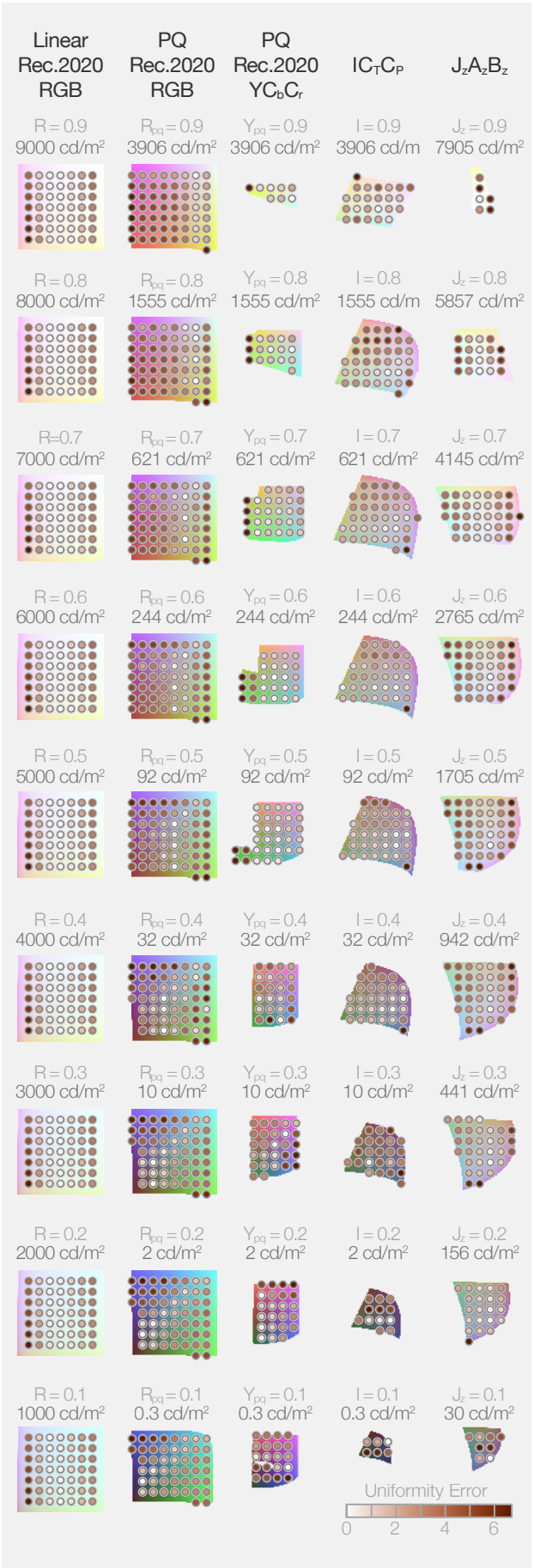


***Figure 8.*** *Perceptual uniformity errors distribution over the tested HDR colour samples.*

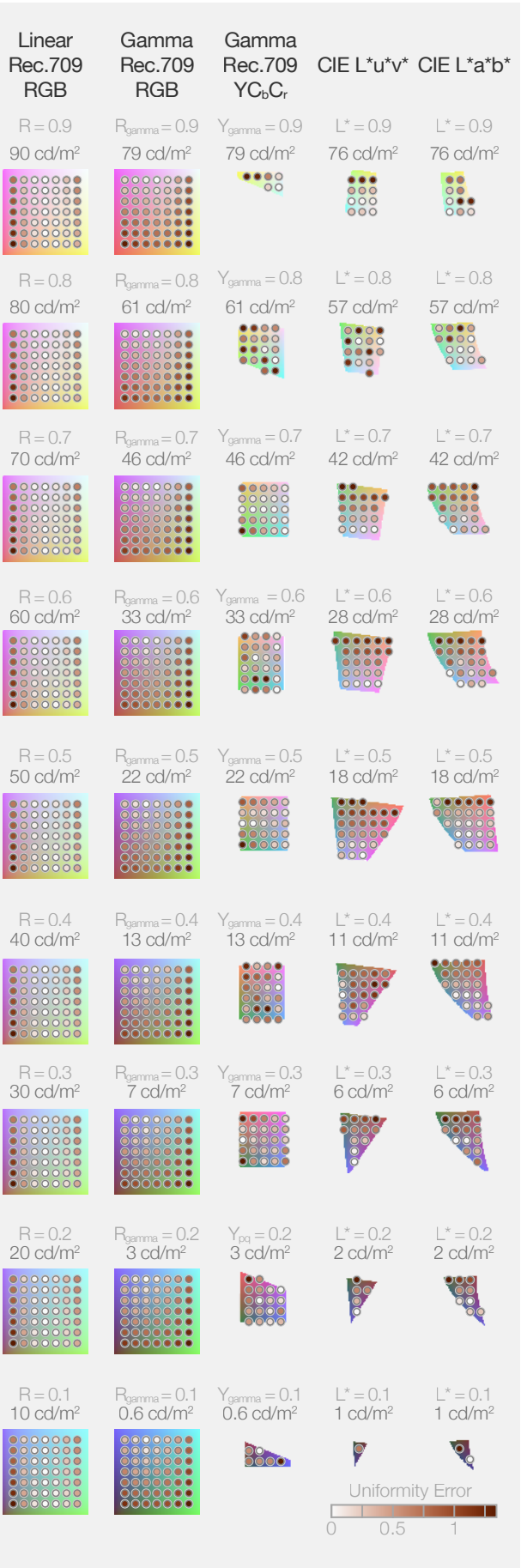


***Figure 9.*** *Perceptual uniformity errors distribution over the tested SDR colour samples.*