https://doi.org/10.2352/ISSN.2470-1173.2018.12.IQSP-299

Advantages of Incorporating Perceptual Component Models into a Machine Learning framework for Prediction of Display Quality

Anustup Choudhury, Scott Daly; Dolby Laboratories Inc.; Sunnyvale, CA, USA

Abstract

Recent work in prediction of overall HDR and WCG display quality has shown that machine learning approaches based on physical measurements performs on par with more advanced perceptually transformed measurements. While combining machine learning with the perceptual transforms did improve over using each technique separately, the improvement was minor. However, that work did not explore how well these models performed when applied to display capabilities outside of the training data set. This new work examines what happens when the machine-learning approaches are used to predict quality outside of the training set, both in terms of extrapolation and interpolation. While doing so, we consider two models – one based on physical display characteristics, and a perceptual model that transforms physical parameters based on human visual system models. We found that the use of the perceptual transforms particularly helps with extrapolation, and without their tempering effects, the machine learning-based models can produce wildly unrealistic quality predictions.

Introduction

High dynamic range (HDR) and wide color gamut (WCG) capability have now become mainstream in consumer TV displays, and is making headway into desktop monitors, laptops, and mobile device products. However, there is not one set of display parameters that define such capability. Rather, there is a continuum of physical ranges generally dependent on cost, and the resulting perceived quality depends non-linearly on those ranges. Being able to quantify the perceived quality for these new and differing physical capabilities has become necessary for the display business.

Various tools in machine learning have increasingly been used to generate quality models based on subjective test data sets, with most activity being in terms of video compression quality, which is signal-dependent [1, 2, 3, 4, 5, 6, 7, 8, 9]. However, display design generally favors signal-independent metrics that can be determined from measurements of the display by a small number of synthetic test images. Signal independent approaches have been used for subjective studies investigating range issues of key HDR parameters, such as contrast and brightness [10], perceived HDR range [11], maximum luminance [12], and backlight modulation [13].

From our experience in developing HDR displays, we think there are five key HDR display parameters: maximum luminance, minimum luminance, local contrast, bit-depth, and color volume. Maximum luminance or 'peak white' of the range is very important for enabling the highlights that distinguishes HDR from SDR. Minimum luminance or 'black level' is also important for achieving the perception of depth that is often described for 2D HDR, as well as purely aesthetic reasons. Local contrast is the technology-neutral term that encompasses the resolution of backlight modulation, and thus the spatial aspects of HDR performance. Bit-depth addresses quantization precision, and the most noticeable distortion from insufficient bit-depth is false edges (contouring, banding). Lastly, color volume encompasses the max luminance, the minimum luminance, and the color gamut.

We studied the overall quality of these five display parameters with subjective tests where each was manipulated through a range going from SDR to one of the most capable HDR displays. For display quality, most existing work uses models with either one display parameter or combines multiple parameters [14, 15] to estimate display quality that correlates with user preference. We consider two models – one based on the physically measured display characteristics and another that transforms those using models of the human visual system (HVS). All existing methods predict quality of displays that are in their training set. In contrast to those works, we explore how well the models predict quality for displays that are outside of the training set. Accordingly, we investigate interpolation and extrapolation capabilities of these models.

Experimental Setup

The underlying data was gathered from a subjective experiment which compared a short video sequence displayed with the best HDR quality available to our lab against the same content shown with reduced display capabilities. The content were all shown on a reference monitor known as the Pulsar, manufactured by Dolby, using dual modulation (a.k.a local dimming). Of the five key display parameters being studied, this display's capabilities are 4000 cd/m^2(nits) maximum luminance, 0.005 cd/m^2 minimum luminance, 12 bits/color in the SMPTE 2084 nonlinear luminance domain, an RGB backlight resolution of 104 x 58 (6032 zones, identical horizontal and vertical aspect ratios), and a P3 color gamut (DCI cinema). The LCD panel was 1920 x 1080 IPS and the frame rate was 24 fps. We provide a high-level overview of the experimental setup in this section. For more details, please refer to our previous works [14, 15].

Stimuli

A total of 27 clips that included content from studio movies and broadcast, optically captured and computer generated, and all graded and mastered at 4000 cd/m^2 maximum luminance, DCI-P3, 12 bits, full resolution, and 0.005 minimum luminance were used. Similar to a previous study [12], the maximum luminance levels and the color gamut areas were tested in a multivariate design. The white point was calibrated to D65. We also probed black levels, bit-depth, and backlight resolution independently, in a uni-variate design. The tested parameter variations are as follows:

- Maximum Luminance: 100, 400, 1000, 4000 cd/m^2
- Color Gamut: BT. R.709, DCI-P3
- Black Level: 0.005, 0.01, 0.05, 0.1 cd/m^2
- Bit Depth: 12, 10, 8, 7 bits
- Backlight Resolution: Pulsar's full backlight resolution; $1/4$ resolution; $1/8$ resolution; global dimming (i.e., no resolution)

The tested variables were compared against a reference sequence, which was always the Pulsar display's best capability.

Viewing Conditions

Single participants viewed the sequences in a dark ambient environment. The viewing distance was three picture heights, giving a field of view (FOV) of approximately 33°. Sequences were presented in a simultaneous, split-screen side-by-side format, randomly presenting the reference image on either the right or left side of the display. Subjects input their quality rating responses via a slider. A GUI showing their rating response was presented on a separate display that was placed below the viewing display.

Task

Participants were instructed to rate each version of the sequence on a continuous -20 to 100 scale. They were instructed that the zero-point should correspond to their memory of SDR quality, and to use 100 for the best HDR quality they have seen. We allowed participants to rate quality below 0 in the instance that they felt a sequence appeared to have a sub-SDR quality. For each trial, one half of the display was the reference display parameters, which served as a hidden upper anchor. They were not informed of the specific distortion applied to the other half.

Modeling Display Characteristics

A general description of the design of a display quality metric is shown in Figure 1. The metric consists of a weighted sum of either physical or perceptually transformed display parameters to form an overall quality metric. Reference and distorted videos are displayed, where in this case the distorted are of a display with reduced capabilities. The subjective experiment (blue lines) generates estimates of the magnitude of overall quality from the both the reference and distorted pairs shown on the HDR display. The reference display with full capabilities serves as an upper anchor. Regression techniques based on machine learning (red lines) is used to tune the weightings of the key display parameters to minimize the difference between the predicted quality rating and the subjective scores.

In order to understand display characteristics, we consider two different models — a physical model that uses physically measured characteristics of the display and a perceptual model that transforms the physical parameters by using HVS-based models. We provide a high-level overview of the models in this section. Please refer to our previous works [14, 15] for more details.

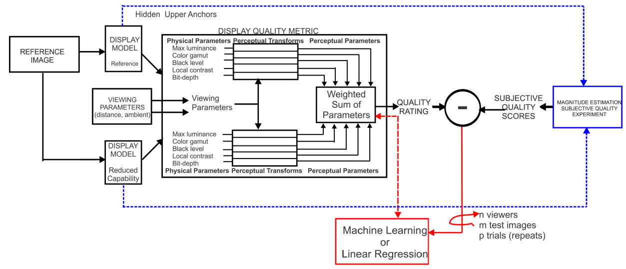


__Figure 1.__ Description of general display quality metric development.

Physical Model

This model uses parameters that can be directly measured. We consider the following 5 parameters:

- Maximum luminance (L_W) – This parameter describes the maximum luminance (peak white) of a display.
- Minimum luminance (L_K) – This parameter describes the minimum luminance (black level) of a display.
- Color gamut (C_G) – The color gamut is determined by measuring the RGB primaries and converted to area in x,y.
- Bit depth (B) – This parameter describes the bit depth of the content that is being presented on the display.
- Backlight resolution (R_B)– We use angular resolution of horizontal zones to calculate backlight resolution. We calculate the angular resolution as follows –
$$R_B = \frac{FOV}{Z_H}, \tag{1}$$
where Z_H is the horizontal number of zones and $FOV = 33°$.

Perceptual Model

The perceptual model is derived by transforming the physical parameters using human vision system (HVS) models:

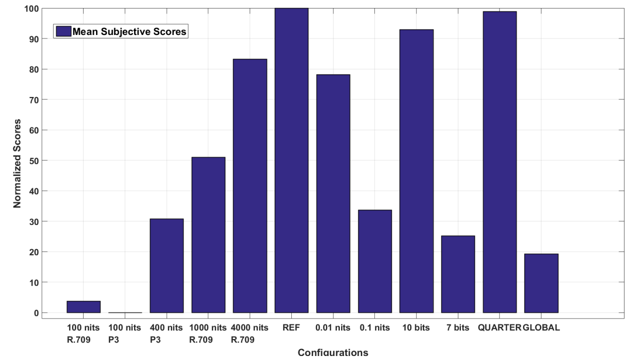
- Maximum luminance (L_{W-HVS}) – This parameter is obtained by applying the SMPTE ST-2084 Perceptual Quantizer (PQ) EOTF transfer function [16] to the L_W parameter of the physical model and is based on the light-adaptive contrast sensitivity function of the human visual system [17, 18, 19, 20].
- Minimum luminance, i.e. black level (L_{K-HVS}) – This parameter is obtained by similarly applying the PQ transfer function to the L_K parameter of the physical model .
- Color Volume (C_{V-HVS}) – This parameter is used to describe the range of colors produced by high dynamic range and wide color gamut displays. It calculates the volume of the 3D color solid in a perceptually uniform space in the IC_TC_P domain.
- Bit depth JND (B_{HVS}) – The perceptual aspect of bit depth was based on computing the number of distinguishable gray (NDG) levels [21], with a small deviation. We calculated this by first converting linear luminance to normalized PQ values. These PQ values are then quantized according to JND experiments. Finally, we computed the maximum difference between two consecutive quantized values.
- Backlight resolution (R_{B-HVS}) – To transform backlight resolution of displays into a perceptual model, we use a contrast metric called Perceptual Contrast Area [14, 15] (PCA) that performs a PSF (point-spread function) analysis of the local contrast capabilities of the display.

Results and Discussion

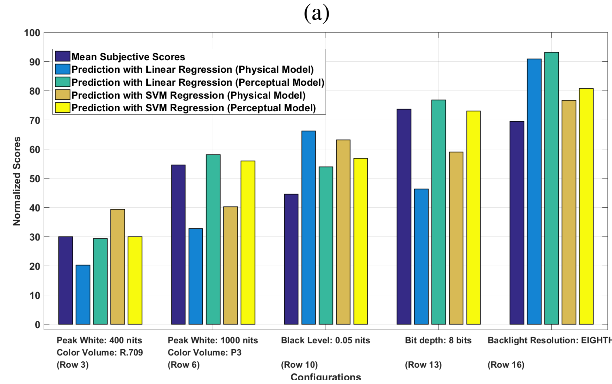
In this section, we illustrate the performance of the physical and perceptual models on data that lie outside the specifications of the training set – i.e., we want to understand the interpolation and extrapolation capabilities of our models. Interpolation is the process of determining values at arbitrary points between two points with known values. On the other hand, extrapolation is the process of determining values at arbitrary points beyond the range that is certainly known. We simulate 17 different display characteristics, as combinations of the parameters mentioned in the previous section and the parameter values of the physical and perceptual model is shown in Table 1.

For each of the 17 display configurations shown in Table 1, we collected subjective scores across all participants. These scores were first normalized to account for intra-participant variations in their range of responses. Finally, we computed the mean of those scores i.e., a mean opinion score (MOS) for each display.

To learn the relationship between models and the subjective scores (MOS), we compared linear regression and machine learning techniques such as Support Vector Machine (SVM) regression [22] and Random Forests [23]. We use an RBF kernel for the SVM [22]. These machine learning methods are used to train and test both physical and perceptual models using the MOS. Since in our previous work [15], we have shown that SVM outperforms both Multilayer perceptron [24] regression and Radial Basis Function (RBF) [25, 26] network regression, we do not use those methods in this analysis. For validation, we use 5-fold cross-validation.



(a)



(b)

Figure 2. *Interpolation capability of our models. (a) MOS of displays used for training (b) Predicted MOS on interpolated test set*

Interpolation

In order to test the interpolation capability of the models, we tested their performance for ranges that lie within the interval between the minimum and maximum value of the training data set. Rather than training on the entire subjective data set, we omitted specific parameters in the training and tested how well that particular trained model could predict the omitted parameter's subjective results. Specifically, we trained the models on rows 1, 2, 4, 5, 7, 8, 9, 11, 12, 14, 15 and 17 and tested it on rows 3, 6, 10, 13 and 16 of Table 1. Since our training set included displays with maximum luminance in $[100, 4000]$ nits range, black levels in $[0.005, 0.1]$ nits range, bit depths in $[7, 12]$ range and backlight resolution in the range between full local dimming and global dimming resolution, we tested the models on displays with the following non-trained configurations – 400 nits maximum luminance and BT. R. 709 color gamut, 1000 nits maximum luminance and P3 color gamut, black level of 0.05 nits, bit depth of 8 bits and backlight resolution of $1/8$. Note that the tested configurations lie within the range of the training data set. Also, the configuration that is being tested is not included during training. We normalized the ratings to give the best display in the training set a score of 100 and the worst display a score of 0 as seen in Figure 2(a). Using those normalization parameters, we get subjective ratings for the test displays. Considering the MOS of the testing set as the "ground truth", we evaluate the performance of each model by comparing its predicted scores using different methods with the "ground truth".

From Figure 2(a) we can see that, the Pulsar 100 nits/ P3 display has a MOS of 0 and the Pulsar reference display has a MOS of 100. Figure 2(b) shows the prediction performance of the models in terms of interpolation capability. We can see that perceptual models are better at interpolation than physical models. Also, SVM is better at prediction than linear regression. We do not visualize the results using Random Forests since it performs worse than SVM (Refer to Table 2).

To quantify the performance, we use two standard performance evaluation procedures and criteria [27] – Root mean square error (RMSE) and Pearson linear correlation coefficient (PLCC). RMSE is used for measuring prediction consistency and PLCC for prediction accuracy. Lower values of RMSE indicates better performance and higher values of PLCC imply better accuracy. Table 2 provides the comparison between the models.

From Table 2, we can confirm that machine learning techniques are generally better at prediction than the simple linear regression method. Also, amongst the machine learning techniques, SVM [22] Regression showed better performance than Random Forests [23]. We can also confirm that the perceptual model is better at prediction than the physical model (e.g., 0.95 Vs 0.61 for PLCC). Combining machine learning techniques with the perceptual model has the best performance. However, its performance is only marginally better than using a perceptual model with simple linear regression method.

Extrapolation

In order to test the extrapolation capability of the models, we tested their performance for ranges that lie beyond the interval between the minimum and maximum value of the training data set. We trained the models on rows 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 15 and 16 and tested on rows 1, 2, 11, 14 and 17 of Table 1. Since

Table 1: Display characteristics for both physical and perceptual model. Key reduced physical parameters are marked in blue. Otherwise, they match the reference capability (marked in yellow). Indices marked in red are used to test interpolation capability and those marked in green are used to test extrapolation capability.

Idx.	Display	Maximum luminance		Minimum luminance		Color volume		Bit depth		Backlight resolution	
		L_W	L_{W-HVS}	L_K	L_{K-HVS}	C_G	C_{V-HVS}	B	B_{HVS}	R_B	R_{B-HVS}
		cd/m^2 (nits)		cd/m^2 (nits)		Area in xy		bits		$^\circ$/zone	
1	100/R709	100	0.5081	0.005	0.0151	0.1120	0.0287	12	0.0002	0.317	0.8546
2	100/P3	100	0.5081	0.005	0.0151	0.1520	0.0405	12	0.0002	0.317	0.8546
3	400/R709	400	0.6526	0.005	0.0151	0.1120	0.0510	12	0.0002	0.317	0.8546
4	400/P3	400	0.6526	0.005	0.0151	0.1520	0.0720	12	0.0002	0.317	0.8546
5	1000/R709	1000	0.7518	0.005	0.0151	0.1120	0.0690	12	0.0002	0.317	0.8546
6	1000/P3	1000	0.7518	0.005	0.0151	0.1520	0.0975	12	0.0002	0.317	0.8546
7	4000/R709	4000	0.9026	0.005	0.0151	0.1120	0.0990	12	0.0002	0.317	0.8546
8	Reference	4000	0.9026	0.005	0.0151	0.1520	0.1400	12	0.0002	0.317	0.8546
9	0.01	4000	0.9026	0.01	0.0215	0.1520	0.1398	12	0.0002	0.317	0.8546
10	0.05	4000	0.9026	0.05	0.0461	0.1520	0.1389	12	0.0002	0.317	0.8546
11	0.1	4000	0.9026	0.1	0.0623	0.1520	0.1382	12	0.0002	0.317	0.8546
12	10 bits	4000	0.9026	0.005	0.0151	0.1520	0.1400	10	0.0039	0.317	0.8546
13	8 bits	4000	0.9026	0.005	0.0151	0.1520	0.1400	8	0.0588	0.317	0.8546
14	7 bits	4000	0.9026	0.005	0.0151	0.1520	0.1400	7	0.2284	0.317	0.8546
15	1/4 Res.	4000	0.9026	0.005	0.0151	0.1520	0.1400	12	0.0002	1.27	0.8544
16	1/8 Res.	4000	0.9026	0.005	0.0151	0.1520	0.1400	12	0.0002	2.54	0.8501
17	Global Res.	4000	0.9026	0.005	0.0151	0.1520	0.1400	12	0.0002	33	0.6194

Table 2: Quantitative comparison of interpolation capability.

	RMSE	PLCC
Physical Model		
Linear Regression	0.1207	0.5533
SVM [22] Regression	0.0768	0.6138
Random Forests [23]	0.1204	0.5967
Perceptual Model		
Linear Regression	0.0660	0.9392
SVM [22] Regression	0.0426	0.9455
Random Forests [23]	0.0655	0.9446

our training set now included displays with maximum luminance in $[400, 4000]$ nits range, black levels in $[0.005, 0.05]$ nits range, bit depths in $[8, 12]$ range and backlight resolution from $[1/8, 1]$ resolution, we tested the models on displays with the following configurations – 100 nits maximum luminance and color gamuts of BT. R. 709 and P3, black level of 0.1 nits, bit depth of 7 bits and global dimming backlight resolution. The tested configurations correspond to endpoints of particular parameter ranges. The chosen parameters for testing all lie on the lower end of the ranges. We did not select parameters that lie on the other end of the spectrum viz., 4000 nits maximum luminance, 0.005 nits black level, bit depth of 12 and full local dimming backlight resolution since those correspond to Pulsar's native display (Reference) and are common to most display configurations. Removing them from training would result in a very small training set, that would not be conducive for learning.

Figure 3(a) shows the MOS of the training set and we can see that, the Pulsar 400 nits/ R.709 display has a MOS of 0 and the Pulsar reference display has a MOS of 100. Figure 3(b) shows the prediction performance of the models in terms of their extrapolation capability. Since our test set contained displays with parameters from the lower end of ranges, we can consider them to have "lower" quality than the ones in the training set. This is illustrated in Figure 3(b) where the MOS of the test set is mostly negative due to normalization parameters from the training set being used to obtain ratings for the test set. From Figure 3(b), we can see that SVM are better at extrapolation than simple linear regression. In general, perceptual models are better at prediction than physical model. Combining linear regression prediction with perceptual model seems to be an exception for predicting global dimming backlight. This is because the variations in R_{B-HVS} in the training set are far less as compared to its value in the test set. In some sense, the value of R_{B-HVS} in the test set seems to be an outlier resulting in bad prediction. However, SVM results in much better prediction in this scenario. Once again, combining perceptual model with SVM (machine learning techniques) are better at prediction than using physical models.

We present quantitative scores of the extrapolation capability of the models in Table 3. Similar to the trends for interpolation, machine learning techniques are better at prediction than linear regression. Also, SVM is better at extrapolation than Random Forests. Perceptual model is also better than physical model. Combining SVM with perceptual model results in best performance. The RMSE for the perceptual model with linear regression is substantially higher than the others because of its bad prediction of global dimming backlight, as seen in Figure 3(b).

For interpolation (Table 2), inclusion of perceptual transforms substantially improves the prediction using SVM. For extrapolation (Table 3), the improvements from using the perceptual transforms is even more substantial than the case for interpolation. As previously mentioned for interpolation, using SVM with perceptual model is marginally better than using simple linear regression with perceptual model. However for extrapolation, using SVM with perceptual model is significantly better than using sim-

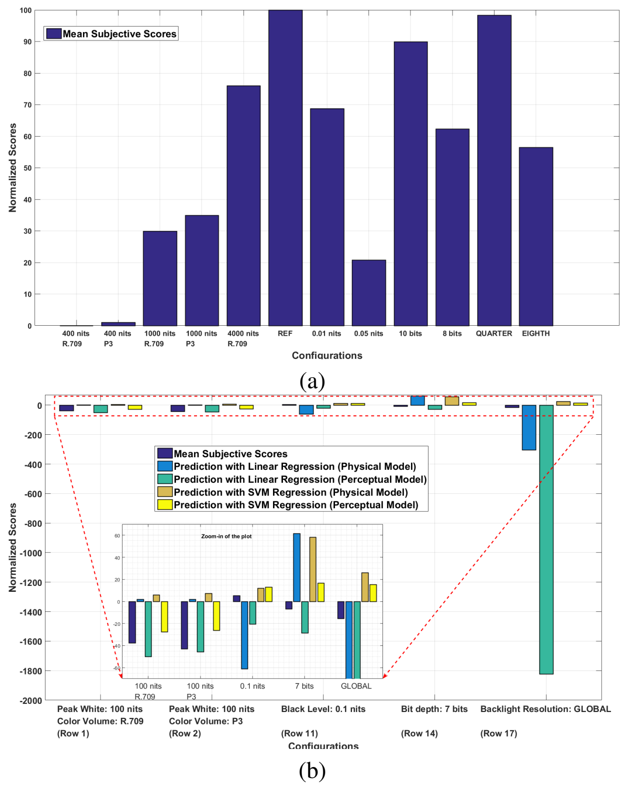


Figure 3. *Extrapolation capability of our models. (a) MOS of displays used for training (b) Predicted MOS on extrapolated test set*

ple linear regression with perceptual model.

Prediction outside our subjective study

We also explored display characteristics outside our subjective study for which we have extraneous evidence about subjective ratings. In order to test such displays, we trained our models on all rows of Table 1 and tested it on two values – .

- Bit-depth of 14 bits
- 1/2 backlight resolution

It is already known from extraneous experiments, that for the PQ signal range of 0–10000 cd/m^2 there is no distortion visibility, and 14 bits would show no advantages [19]. For backlight resolution of 1/2, our reference pilot studies showed the quality was identical to the reference backlight resolution for the three picture heights viewing distance. In order to predict MOS for 14 bits, we explore the extrapolation capability of the models. Likewise, to predict MOS for 1/2 backlight resolution, we explore the interpolation capability of the models.

Table 3: Quantitative comparison of extrapolation capability.

	RMSE	PLCC
Physical Model		
Linear Regression	0.5504	-0.1774
SVM [22] Regression	0.1821	0.4838
Random Forests [23]	0.2854	0.3145
Perceptual Model		
Linear Regression	3.228	-0.0992
SVM [22] Regression	0.0784	0.9114
Random Forests [23]	0.2472	0.7216

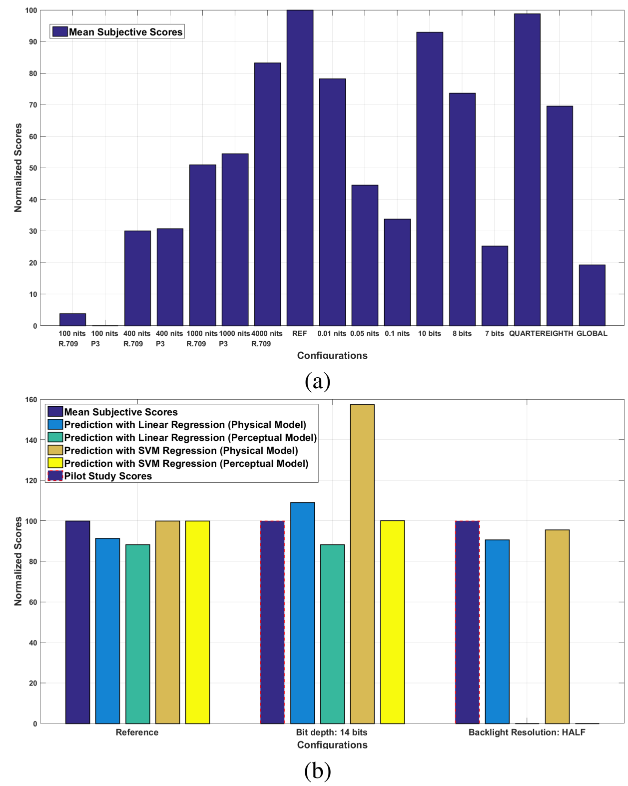


Figure 4. *Prediction for data outside of subjective study. (a) MOS of displays used for training (b) Predicted MOS on test set*

We normalized the ratings to give the best display in the training set a score of 100 and the worst display a score of 0 as seen in Figure 4(a). Using those normalization parameters, we get ratings for the test displays. From Figure 4(a) we can see that, the Pulsar 100 nits/ P3 display has a MOS of 0 and the Pulsar reference display has a MOS of 100. Figure 4(b) shows the prediction results. For the reference display, which is a part of the training set, we can see that SVM has almost perfect prediction, irrespective of the models being used. When extrapolating to 14 bits, combining SVM with perceptual model also has perfect prediction. Surprisingly, using a physical model with linear regression is better than using it with SVM. Also, SVM has better prediction when interpolating to 1/2 backlight resolution compared to linear regression[1]. In general, combining perceptual models with machine learning has the best prediction.

Conclusion & Future Work

In this paper, we test HDR display characteristics and transform that into a single number pertaining to overall subjective quality. This is one of the first attempts at predicting quality of HDR displays that are outside of the training set – both in terms of interpolation and extrapolation. We consider two different models – a physical model and a perceptual model that transforms the physical characteristics using a HVS model. In addition to linear regression, we use machine learning techniques such as Random

[1]We don't have measured subjective values for backlight resolution (1/2) for the perceptual model, but we informally know it should be very close to the reference

forests and SVM regression to learn the relationship between the display parameters and the subjective scores. We conclude that a perceptual model is much better at predicting subjective quality than a physical model. Machine learning techniques result in better fit to the data as compared to linear regression. We found that the machine learning approaches are subject to failure cases when tested on conditions outside of their training set. Incorporating perceptually transformed components into the machine learning framework can reduce those failure cases. These effects are more pronounced during extrapolation than during interpolation. Therefore, using machine learning with the perceptual model results in the best performance.

Future work includes ascertaining the significance of these results by conducting more experiments, involving more test images, subjects, and displays. We suspect our test content did not adequately probe the value of extended color gamuts, black levels, or higher bit depths.

References

[1] Mittal, A., Moorthy, A. K., and Bovik, A. C., "No-reference image quality assessment in the spatial domain," *IEEE Transactions on Image Processing* **21**, 4695–4708 (Dec 2012).

[2] Kang, L., Ye, P., Li, Y., and Doermann, D., "Convolutional neural networks for no-reference image quality assessment," in [*2014 IEEE Conference on Computer Vision and Pattern Recognition*], 1733–1740 (June 2014).

[3] Zuo, L., Wang, H., and Fu, J., "Screen content image quality assessment via convolutional neural network," in [*2016 IEEE International Conference on Image Processing (ICIP)*], 2082–2086 (Sept 2016).

[4] Li, Z., Aaron, A., Katsavounidis, I., Moorthy, A., and Manohara, M., "Toward a practical perceptual video quality metric." `http://techblog.netflix.com/2016/06/toward-practical-perceptual-video.html/` (2016).

[5] Li, Z., Norkin, A., and Aaron, A., "VMAF - video quality metric alternative to PSNR," *Joint Video Exploration Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11* (October 2016).

[6] Xu, L., Lin, W., and Kuo, C.-C. J., [*Visual Quality Assessment by Machine Learning*], Springer Publishing Company, Incorporated (2015).

[7] Ali Amirshahi, S., Pedersen, M., and Yu, S. X., "Image quality assessment by comparing cnn features between images," *Electronic Imaging* **2017**(12), 42–51 (2017).

[8] Alam, M. M., Patil, P., Hagan, M. T., and Chandler, D. M., "A computational model for predicting local distortion visibility via convolutional neural network trained on natural scenes," in [*2015 IEEE International Conference on Image Processing (ICIP)*], 3967–3971 (2015).

[9] Sheikh, H. R. and Bovik, A. C., "A visual information fidelity approach to video quality assessment," in [*First International Workshop on Video Processing and Quality Metrics for Consumer Electronics*], 23–25 (2005).

[10] Seetzen, H., Li, H., Ye, L., Heidrich, W., Whitehead, L., and Ward, G., "25.3: Observations of luminance, contrast and amplitude resolution of displays," *SID Symposium Digest of Technical Papers* **37**(1), 1229–1233 (2006).

[11] Hulusic, V., Valenzise, G., Provenzi, E., Debattista, K., and Dufaux, F., "Perceived dynamic range of hdr images," in [*2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX)*], 1–6 (June 2016).

[12] Hanhart, P., Korshunov, P., Ebrahimi, T., Thomas, Y., and Hoffmann, H., "Subjective quality evaluation of high dynamic range video and display for future tv," *SMPTE Motion Imaging Journal* **124**, 1–6 (May 2015).

[13] Mantel, C., Korhonen, J., Forchhammer, S., Pedersen, J., and Bech, S., "Subjective quality of videos displayed with local backlight dimming at different peak white and ambient light levels," in [*2015 Seventh International Workshop on Quality of Multimedia Experience (QoMEX)*], 1–6 (May 2015).

[14] Choudhury, A., Farrell, S., Atkins, R., and Daly, S., "55-1: Invited paper: Prediction of overall hdr quality by using perceptually transformed display measurements," *SID Symposium Digest of Technical Papers* **48**(1), 819–822 (2017).

[15] Choudhury, A., Farrell, S., Atkins, R., and Daly, S., "Prediction of hdr quality by combining perceptually transformed display measurements with machine learning," in [*Proc. SPIE*], **10396**, 10396 – 10396 – 16 (2017).

[16] "ST 2084:2014 - SMPTE Standard - high dynamic range electro-optical transfer function of mastering reference displays," *SMPTE ST 2084:2014* , 1–14 (Aug 2014).

[17] Cowan, M., Kennel, G., Maier, T., and Walker, B., "Contrast sensitivity experiment to determine the bit depth for digital cinema," *SMPTE Motion Imaging Journal* **113**, 281–292 (Sept 2004).

[18] Aydin, T. O., Mantiuk, R., and Seidel, H.-P., "Extending quality metrics to full luminance range images," *Proc. SPIE* **6806**, 68060B–68060B–10 (2008).

[19] Miller, S., Nezamabadi, M., and Daly, S., "Perceptual signal coding for more efficient usage of bit codes," in [*The 2012 Annual Technical Conference Exhibition*], 1–9 (Oct 2012).

[20] Nezamabadi, M., Miller, S., Daly, S., and Atkins, R., "Color signal encoding for high dynamic range and wide color gamut based on human perception," *Proc. SPIE* **9015**, 90150C–90150C–12 (2014).

[21] Ward, G., "59.2: Defining dynamic range," *SID Symposium Digest of Technical Papers* **39**, 900–902 (2008).

[22] Cortes, C. and Vapnik, V., "Support-vector networks," *Machine Learning* **20**, 273–297 (Sept. 1995).

[23] Breiman, L., "Random forests," *Machine Learning* **45**, 5–32 (Oct 2001).

[24] Rumelhart, D. E., Hinton, G. E., and Williams, R. J. in [*Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1*], ch. Learning Internal Representations by Error Propagation, 318–362, MIT Press, Cambridge, MA, USA (1986).

[25] Orr, M. J. L., "Introduction to radial basis function networks," (1996).

[26] Wettschereck, D. and Dietterich, T. G., "Improving the performance of radial basis function networks by learning center locations," in [*NIPS*], **4**, 1133–1140, Morgan Kaufmann (1991).

[27] VQEG, "Final report from the video quality experts group on the validation of objective models of video quality assessment." `http://www.its.bldrdoc.gov/vqeg/vqeg-home.aspx/` (2003).