# Supervised Learning in Automatic Selection of Preferred Inverse Tone-Mapping Operator for HDR Display

**Hsuan-Chi Huang[1], Pei-Li Sun[1], Hung-Chung Li[2]**

plsun@mail.ntust.edu.tw
[1]National Taiwan University of Science and Technology, Taipei 10607, Taiwan
[2]Research Center for Information Technology Innovation, Academia Sinica, Taipei 115024, Taiwan
Keywords: High dynamic range, HDR display, Inverse tone-mapping, Machine learning, Deep learning

## ABSTRACT

*Preferred SDR-to-HDR images for HDR display depend on not only inverse tone-mapping methods, but also image contents. We build an operation prediction system based on supervised learning like machine learning and deep learning. A validation experiment demonstrates that our prediction system is highly reliable.*

## 1 INTRODUCTION

The technology and market of high dynamic range (HDR) display has become matured, but the content sources have not kept up with its development. This makes HDR displays can only display standard dynamic range (SDR) images most of the time, and don't make good use of the features and advantages of wide dynamic range.

## 2 RELATED WORKS

### 2.1 Inverse Tone-mapping

Inverse tone-mapping operators (ITMO) could expand the dynamic range of SDR images into high dynamic range, urging the mainstream SDR images to be converted into HDR images immediately. The HDR images utilize the full dynamic range of HDR displays, then further improve the quality of imaging.

However, the existing ITMOs were designed based on different concepts and purposes. For example, based on expansion methods, the ITMOs could be classified into 3 different types [1]. Global models [2][3] apply the same tone expansion function on each pixel in a SDR image. Classification models [4][5] segment a SDR image into different parts, then each part is expanded by the appropriate function. Expand-map models [6][7][8][9] use an expand-map to guide a SDR image to expand the luminance of each pixel.

Depending on image contents, there might be different optimal ITMOs for the SDR images. There still no consensus which is the best operator for which task [1].

### 2.2 Artificial intelligence

This study attempted to create an optimal ITMO prediction system based on supervised learning, the system (referring to **Fig.1**) was trained by artificial intelligence (AI) algorithms like machine learning (ML) and convolutional neural network (CNN), which are good at image content analysis.

For the ML algorithms, we considered Logistic regression, Support vector machine (SVM), K-nearest neighbors, Naive Bayes classifier, and Multilayer perceptron (MLP); while for the CNN architectures, we used LeNet [10], AlexNet [11], ZFNet [12], VGG16 [13], and ResNet-50 [14].

## 3 EXPERIMENTS

### 3.1 Labeling Experiment

Sasaki [15] conducted a labeling experiment to connect the most preferred tone-mapping operator

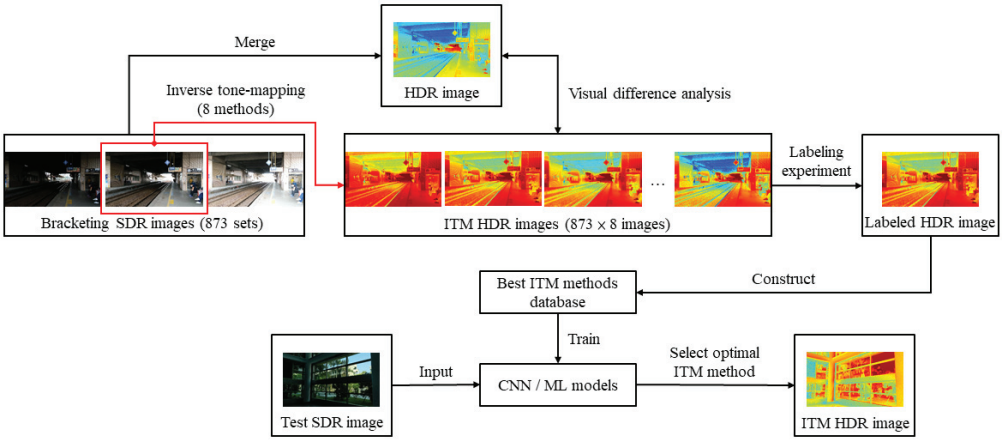


**Fig. 1 Overview of the proposed method.**

(TMO) with image content. On the contrary, we conducted a labeling experiment but to connect the preferred ITMOs with image content, then yielded a database of correspondence. There were 3 subjects (2 females, 1 males) participated in our labeling experiment, each subject labeled 295 sets of HDR images, a total of 873 sets of HDR images prepared for this labeling experiment (6 for repeatability test).

**Fig. 2 Experimental setup: (top) viewing environment, (bottom) experiment GUI.**

The labeling experiment was performed in a dark room to fully utilize the dynamic range of display (referring to **Fig.2**). Stimuli were shown on a VESA certificated HDR display, the max luminance of which achieved 400 cd/m$^2$ in HDR mode, and all the inverse tone-mapping HDR images were encoded by Perceptual Quantization (PQ) [16]. The PQ character of the HDR display is shown in **Fig. 3**.

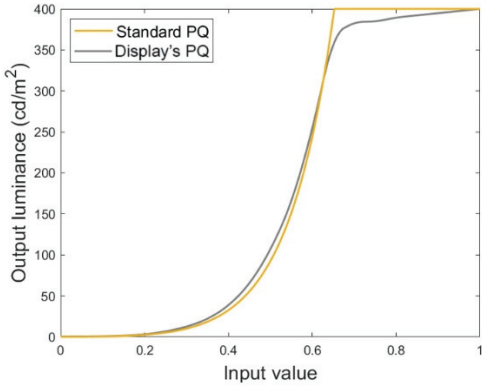


**Fig. 3 PQ character of experiment HDR display.**

When the experiment was proceeding, a SDR image was first shown on a calibrated SDR display (max luminance 100 cd/m$^2$ according to BT.1886 [17]) on the right side of **Fig.2** for 3 seconds, for subject to observe the original appearance of the image. Afterwards, the ITM HDR images expanded from the exact right SDR image by 8 different ITM methods were shown on the HDR display ahead simultaneously and randomly on the left side **o**f **Fig.2**. Then the subject labeled his/her most preferred ITM HDR images in this set. The labelling results are summarized in **Fig.4.** Thereby we could get a database of correspondence between image contents and preferred ITMOs.

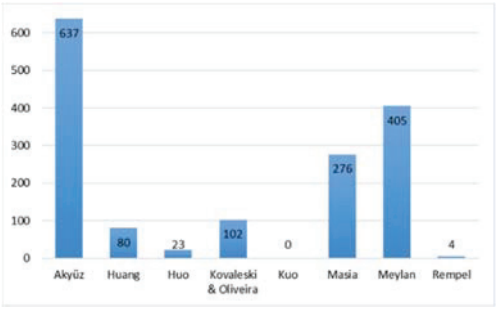


**Fig. 4 Statistical results of labeling experiment.**

This database was trained by ML algorithms and CNN architectures mentioned before, the test accuracy (independent from training set) achieved about 76% for CNN and 78% for ML.

### 3.2 Validation Experiment

A human-factor validation experiment was conducted to investigate the performance of the trained prediction systems in practice.

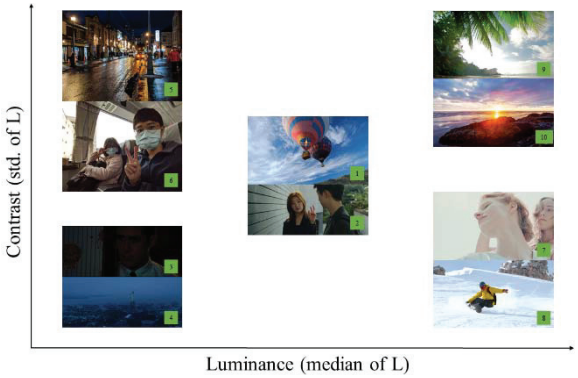


**Fig. 5 10 images used in validation experiment.**

Independent 10 SDR images with different lighting style classifications based on [18] shown in **Fig.5** were prepared and 15 subjects (5 females, 10 males) participated in this validation experiment. The procedure of validation experiment was similar to the labeling experiment, but was distinct from the labeling part. In the validation experiment, the subject wouldn't label the most preferred ITM HDR images, but evaluated every ITM HDR images on a preference scale from 1 (very bad) to 7 (very good) as shown as **Fig.6**. The visual scoring results were analyzed by Analysis of Variance (ANOVA), and Dunnett's T3 test was used to determine 95% confidence intervals of each ITMO.

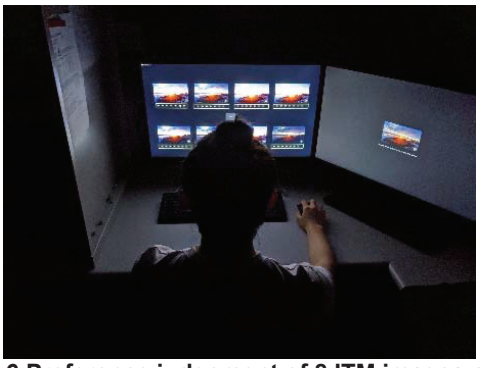

**Fig. 6 Preference judgement of 8 ITM images on the left HDR display (according to its SDR original shown on the right side) in the validation experiment.**

## 4 RESULTS

### 4.1 Prediction Systems

The 10 SDR images in the validation experiment were respectively analyzed by our trained systems to predict which ITMO may be the most optimal for this image to expand to HDR image. Briefly compare the prediction of our systems and the result scores of the validation experiment, both AlexNet and VGG16 models trained with gray images (Y of YCbCr) got the highest accuracy, and the general accuracy fell between 40% to 50%. On the other hand, classifier chain with MLP got the lowest accuracy where the mean accuracy is close to 33%. **Table 1** lists the preferred ITMO choice by 6 different AI models. The green cells represent the choice match the visual experimental results, whereas the cells with white background mismatch the visual results.

**Table 1 ITMO predictions of 6 trained models. Green cells represent perfect match to the results of the validation experiment.**

| Image | Classifier chain + MLP (10,10) | Label powerset + MLP (10,10) | AlexNet (RGB) | ZFNet (RGB) | AlexNet (Y) | VGG16 (Y) |
|---|---|---|---|---|---|---|
| 1 | Masia | Masia | Masia | Masia | Masia | Akyüz |
| 2 | Masia | Masia | Masia | Masia | Akyüz | Akyüz |
| 3 | Akyüz | Akyüz | Akyüz | Akyüz | Akyüz | Akyüz |
| 4 | Meylan | Masia | Masia | Masia | Masia | Akyüz |
| 5 | Akyüz | Akyüz | Akyüz | Akyüz | Akyüz | Akyüz |
| 6 | Masia | Masia | Masia | Masia | Akyüz | Akyüz |
| 7 | Masia | Masia | Masia | Masia | Masia | Akyüz |
| 8 | Masia | Masia | Masia | Masia | Masia | Akyüz |
| 9 | Masia | Masia | Masia | Masia | Masia | Akyüz |
| 10 | Akyüz | Akyüz | Akyüz | Akyüz | Akyüz | Akyüz |

However, if we take confidence intervals of the validation experiment into account, the prediction got nearly 100% accurate, except the “classifier chain with MLP” mis-predicted the preferred ITMO for No.4 image. The blue cells in **Table 2** represent the AI predictions which are not the best choices in the validation experiment but the predictions have no significant differences to the best choices on the preference scores statistically. It demonstrates that our prediction systems are highly reliable.

**Table 2 Blue cells represent the ITMO predictions have insignificant differences (acceptable match) to the results of the validation experiment.**

| Image | Classifier chain + MLP (10,10) | Label powerset + MLP (10,10) | AlexNet (RGB) | ZFNet (RGB) | AlexNet (Y) | VGG16 (Y) |
|---|---|---|---|---|---|---|
| 1 | Masia | Masia | Masia | Masia | Masia | Akyüz |
| 2 | Masia | Masia | Masia | Masia | Akyüz | Akyüz |
| 3 | Akyüz | Akyüz | Akyüz | Akyüz | Akyüz | Akyüz |
| 4 | Meylan | Masia | Masia | Masia | Masia | Akyüz |
| 5 | Akyüz | Akyüz | Akyüz | Akyüz | Akyüz | Akyüz |
| 6 | Masia | Masia | Masia | Masia | Akyüz | Akyüz |
| 7 | Masia | Masia | Masia | Masia | Masia | Akyüz |
| 8 | Masia | Masia | Masia | Masia | Masia | Akyüz |
| 9 | Masia | Masia | Masia | Masia | Masia | Akyüz |
| 10 | Akyüz | Akyüz | Akyüz | Akyüz | Akyüz | Akyüz |

### 4.2 Visual Difference Analysis

We also used HDR-VDP-2 [19] to conduct a visual difference analysis to explore the correlation between observers’ preferred ITM HDR rendering from a single SDR image and ground-truth HDR rendering generated by exposure bracketing in ±2EV stops. **Fig.7** shows its workflow.

The HDR images yielded by Kovaleski & Oliveira’s [7] ITMO and Kuo’s [9] ITMO owned the first and the second smallest visual difference from the ground-truth HDR images, but were respectively only the fourth and the last preferable in the labeling experiment.

This implies that those ITMO HDRs having smaller visual difference to the typical exposure bracketing HDR rendering was not preferred in the observers’ judgement. Akyüz’s ITMO had a balanced performance between the two, therefore the ITMO is recommended.

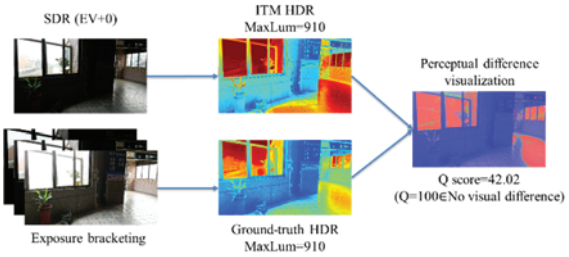


**Fig. 7 Workflow of visual difference analysis**

## 5 CONCLUSIONS

Overall, the AlexNet model trained with gray images is the most recommended prediction system in this study because of its robustness. Akyüz’s ITMO is preferred for most of test images. And the results of visual difference analysis show that preferred ITM HDR rendering is different to the typical exposure bracketing HDR rendering. It suggests that photometrically-accurate HDR rendering doesn’t ensure higher preference in observers’ judgment.

## REFERENCES


[1] F. Banterle, A. Artusi, K. Debattista, and A. Chalmers, *Advanced High Dynamic Range Imaging*, 2nd. ed., CRC Press, 2017.

[2] A. O. Akyüz, R. Fleming, B. E. Riecke, E. Reinhard, and H. H. Bülthoff, "Do HDR displays support LDR content?," *ACM Transactions on Graphics*, vol. 26, no. 99, p. 38, 2007, doi: 10.1145/1239451.1239489.

[3] B. Masia, A. Serrano, and D. Gutierrez, "Dynamic range expansion based on image statistics," *Multimedia Tools and Applications*, vol. 76, no. 1, pp. 631–648, 2017, doi: 10.1007/s11042-015-3036-0.

[4] L. Meylan, S. Daly, and S. Süsstrunk, "Tone mapping for high dynamic range displays," in *Human Vision and Electronic Imaging XII*, Feb. 2007, vol. 6492, p. 649210, 2007, doi: 10.1117/12.706472.

[5] Y.-Z. Lai and P.-L. Sun, "Preferred LDR to HDR Image Conversion for HDR Displays," IDW'17, Sendai, Japan, 2017.

[6] A. G. Rempel and M. Trentacoste, "Ldr2Hdr," *ACM Transactions on Graphics*, vol. 26, no. 99, p. 39, 2007, doi: 10.1145/1239451.1239490.

[7] R. P. Kovaleski and M. M. Oliveira, "High-quality reverse tone mapping for a wide range of exposures," *Brazilian Symposium of Computer Graphic and Image Processing*, pp. 49–56, 2014, doi: 10.1109/SIBGRAPI.2014.29.

[8] Y. Huo, F. Yang, L. Dong, and V. Brost, "Physiological inverse tone mapping based on retina response," *The Visual Computer*, vol. 30, no. 5, pp. 507–517, May 2014, doi: 10.1007/s00371-013-0875-4.

[9] P.-H. Kuo, C.-S. Tang, and S.-Y. Chien, "Content-adaptive inverse tone mapping," in *2012 Visual Communications and Image Processing*, vol. 23, no. 3, pp. 1–6, 2012, doi: 10.1109/VCIP.2012.6410798.

[10] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998, doi: 10.1109/5.726791.

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017, doi: 10.1145/3065386.

[12] M. D. Zeiler and R. Fergus, "Visualizing and Understanding Convolutional Networks," 2014, pp. 818–833.

[13] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *in 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, pp. 1–14, 2015.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-Decem, no. 50, pp. 770–778, 2016, doi: 10.1109/CVPR.2016.90.

[15] H. Sasaki, K. Hirai, T. Horiuchi, "Automatic Selection of Preferable Tone-Mapping Method Based on Deep Learning," IDW'19, Sapporo, Japan, 2019.

[16] Rec. ITU-R BT. 2100-1, Image parameter values for high dynamic range television for use in production and international programme exchange, 2017.

[17] Rec. ITU-R BT. 1886, Reference electro-optical transfer function for flat panel displays used in HDTV studio production, 2011.

[18] C. Bist, R. Cozot, G. Madec, G. and X. Ducloux, "Style Aware Tone Expansion for HDR Displays," In *Graphics Interface*, pp. 57-63, 2016.

[19] R. Mantiuk, K. J. Kim, A. G. Rempel, and W. Heidrich, "HDR-VDP-2," *ACM Transactions on Graphics*, vol. 30, no. 4, p. 1, 2011, doi:10.1145/2010324.1964935.