

# Automatic Selection of Preferable Tone-Mapping Method based on Deep Learning

Hirofumi Sasaki<sup>1</sup>, Keita Hirai<sup>1</sup>, Takahiko Horiuchi<sup>1</sup>

<sup>1</sup> Chiba University, 1-33, Yayoi-cho, Inage-ku, Chiba-shi, Chiba 263-8522, Japan

Keywords : HDR image database, Preference, Subjective evaluation, Convolutional neural network

## ABSTRACT

*Preference of a tone-mapped HDR image appearance depends on not only an applied tone-mapping method but also an input scene content. In this paper, based on a deep learning technique, we proposed a system to automatically select a tone-mapping method that provides a preferable appearance of an input HDR image.*

## 1 INTRODUCTION

HDR (high dynamic range) images are synthesized by merging images which are taken at different exposures. As a result, users can obtain high-quality images that avoid over-exposure and under-exposure. In order to reproduce HDR images on a standard display, it is necessary to tone-map HDR images and convert them into SDR / LDR (standard / low dynamic range) images.

A lot of tone-mapping methods have been proposed based on various design concepts. Fig.1 shows examples of tone-mapped images by 24 methods. It is known that the preferable appearance of a tone-mapped HDR image depends on not only tone-mapping methods but also scene contents. Therefore, users need to check the processing results which are reproduced by several tone-mapping methods. Finally, it takes time and effort for the users in this selection process of the best tone-mapping method.

In this paper, we propose a system to select the best tone-mapping method based on a deep learning technique such as CNN (convolutional neural network). For achieving this goal, we develop the selection system of a tone-mapping method that provides the most preferable appearance for an input HDR image.

Fig.2 shows overview of this study. For applying deep learning, it is necessary to prepare an HDR image database with preferable score tags. It is well known that the accuracy of a learning model in deep learning techniques significantly depends on a training dataset. However, there are few image datasets tagged with human subjective evaluations. Therefore, first, we construct an HDR image database with preferable score tags through large-scale human subjective evaluation experiments in terms of the preferable appearance of tone-mapped images.

In our subjective experiments, we applied 24 types of tone-mapping methods to 1,996 HDR images collected from the web. In summary, 47,904 tone-mapped images were used as experimental stimuli. In the evaluation



Fig.1 Examples of Tone-Mapped images using 24 methods.

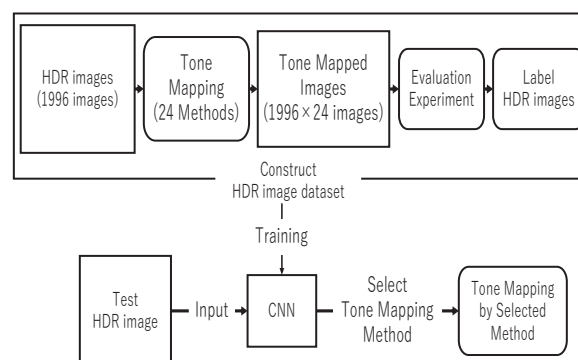


Fig.2 Overview for constructing our system

experiment, the subject evaluated tone-mapped images according to the criteria of “the most preferable appearance.” Thus, a tone-mapping method that provides the most preferable appearance for an input HDR image was determined. Then, we constructed the HDR image database with preferable tags.

Next, we developed a CNN-based system that automatically selected the tone-mapping method that provides the most preferable appearance for an HDR image input. The training data and test data in the CNN were HDR images with preferable tags through the subjective evaluation experiments.

## 2 HDR IMAGE DATASET

This section describes the details of our dataset. First, we collected HDR images from web. Next, we tone-

mapped the HDR images which were used in our subjective experiments.

### 2.1 HDR image collection

The HDR images used in this study were collected from previous studies that were related with HDR images, and websites that distributed free HDR images. Fig.3 shows examples of collected source images. Source HDR images include various real-world scenes with both artificial and natural objects such as mountains, plants, forests, seas, buildings, outdoors, and indoors.

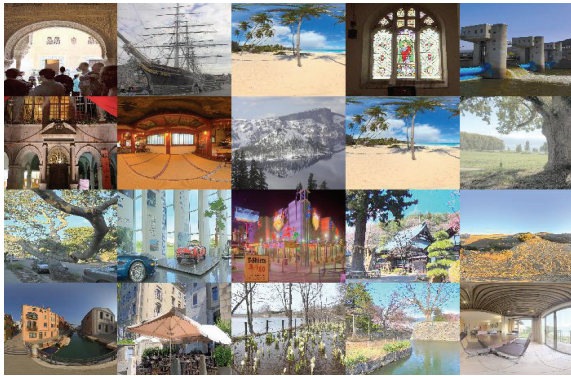


Fig.3 Sample images of our HDR image dataset

### 2.2 Tone-Mapping process

1996 HDR images were used in this study. These HDR images were tone-mapped by 24 types of tone-mapping methods. The tone mapped images were used as experimental stimuli. 24 types of tone-mapping methods are introduced in "High Dynamic Range Imaging" [1]. We applied default parameters to all of tone-mapping methods.

## 3 SUBJECTIVE EXPERIMENT ON PREFERENCE

A subjective evaluation experiment was conducted to determine the tone-mapping method that provides the most preferable appearance for each HDR image.

### 3.1 Stimulus Images

A total of 47,904 stimulus images were prepared by applying 24 types of tone-mapping methods to 1,996 HDR images. Subjects selected an image with most preferable appearance from the 24 tone-mapped images. Thus, 1996 preferable evaluations were conducted in total.

### 3.2 Experiment Environment and Setting

The purpose of our experiments is not to investigate visual characteristics and mechanisms but to construct image database for deep learning. For this reason, we assumed the viewing environment conditions such as illumination light and viewing distance as a general display viewing environment. The experimental environment set up is as follows.

- Display : ASUS PA248Q (sRGB gamut)
- Number of subjects : 17 (15 males and 2 females)
- Number of subjects in each preferable evaluation (24 tone-mapped images) : 1

### 3.3 Experimental Procedure

We instructed the subject to select *"the image that you feel most preferable appearance among the displayed images"*. Next, we explain the evaluation procedure (1 set evaluation).

- step1) 8 images are randomly displayed from 24 types of tone-mapped images. A subject selects the most preferable image.
- step2) 7 images from the unevaluated 16 images (selected randomly) and the image selected in step1 are displayed (8 images in total). The subject selects one image that he feels most preferable again.
- step3) 7 images from the unevaluated 9 images (selected randomly) and the image selected in step2 are displayed (8 images in total). The subject selects one image that he feels most preferable again.
- step4) Following 8 images are displayed. The subject selects the most preferable image.
  - 1 image selected in step 3
  - 2 unevaluated images
  - Image selected in steps 1 and 2 (if different from the image selected in step 3)
  - Randomly select images from the evaluated images for displaying 8 images in total.
- step5) Check the reproducibility. The image selected in step4 and 7 images selected randomly from the other 23 images are displayed. The subject selected the most preferable image.

Finally, the tone-mapping method selected in step 4 is the best method that provides the most preferable appearance for an HDR image input. The tone-mapping method in step 5 is labeled as the following case: a tone-mapped image of the same method is selected in step 1 to step 3 and step 5, and a tone-mapped image that is different from the other steps is selected only in step 4. Fig.4 shows the user interface in the evaluation experiment.

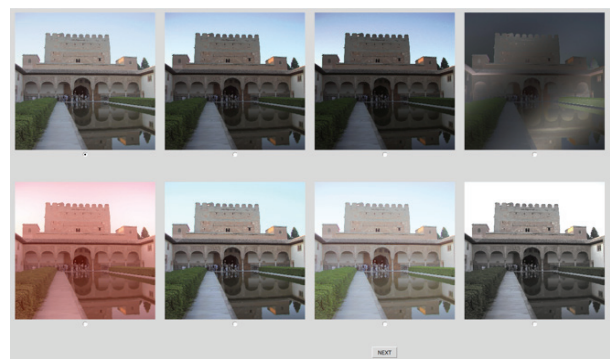
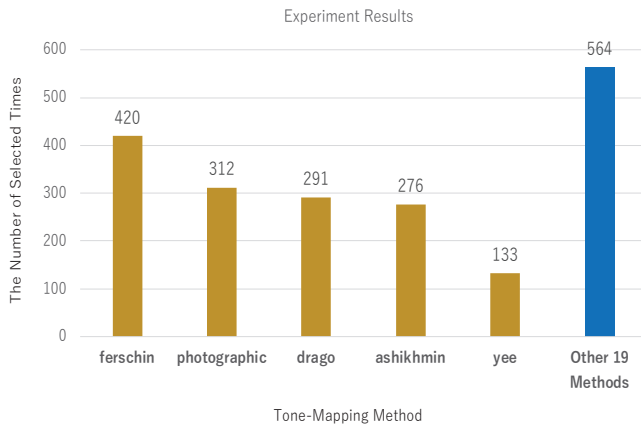


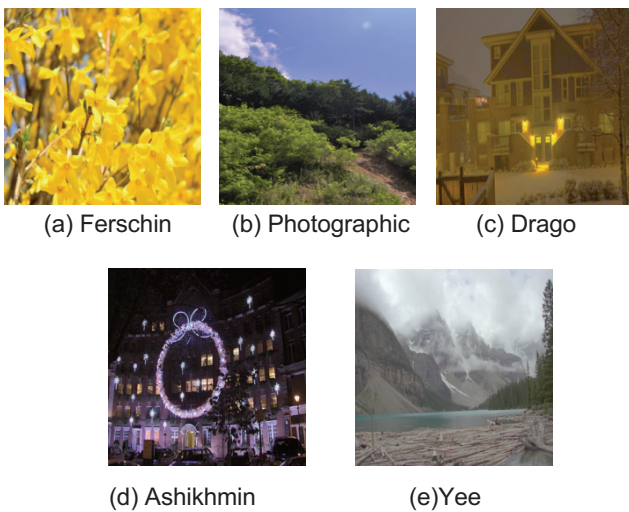
Fig.4 The interface of evaluation experiment

### 3.4 Results

Fig.5 shows the results of the evaluation experiment. The horizontal axis of the graph in Fig.5 indicates each



**Fig. 5 Results of the subjective evaluation experiment**



**Fig.6 The examples of representative images in the top five classes.**

tone-mapping method, and the vertical axis indicates the number of times selected as a method that provides a preferable appearance. In the experimental results, Ferschin [2], Photographic [3], Drago [4], Ashikhmin [5], and Yee [6] were frequently selected.

#### 4 AUTOMATIC TONE-MAPPING METHOD SELECTION USING CNN

##### 4.1 Pre-Processing

Based on the results of the evaluation experiment in Section 3, each HDR image was labeled with the tone-mapping method that provides the most preferable appearance. Based on the experimental results, we constructed an HDR image database. The database is applied to the training in CNN for predicting the most preferable tone-mapping method.

However, as a result of experiments, it is considered difficult to classify all classes because the number of images in each class is different. Therefore, in this study, HDR images of the top five methods in Fig.5 (1,432

images in total: Ferschin [2], Photographic [3], Drago [4], Ashikhmin [5], and Yee [6]) were targeted for classification. The learning model was VGG16 [7], and the number of classes was five. The structure of the network is basically the same as VGG16. Table.1 shows the breakdown of the training data and the test data in our dataset. Fig.6 shows examples of representative images in the class of the top five tone-mapping methods.

##### 4.2 Classification Results

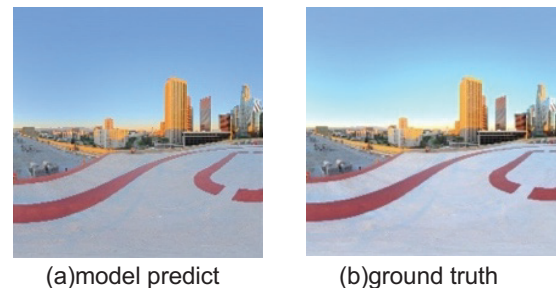
The results of our model prediction was 0.37 (17 correct predictions in 45 test images). The correct answer rate with the second-highest cumulative probability was 0.556 (25 proper predictions in 45 test images).

**Table.1 The breakdown of our dataset**

TM Method	Training data	Test data
Ferschin	410	10
Photographic	302	10
Drago	281	10
Ashikhmin	266	10
Yee	128	5
In total	1,387	45

#### 5 ACCURACY VERIFICATION OF OUR MODEL

Fig.7 shows an example of the result that are the ground truth and the prediction by our model. In this case, the predicted tone-mapping method was different from the ground-truth tone-mapping method which was selected in the subjective evaluation experiment. As shown in this example, the predicted TM method and the ground-truth TM method have almost the same impression and appearance. Even though the model fails the tone-mapping prediction, some predicted method can provide more preferable appearance than the ground-truth method. In this case, our CNN-based prediction model will be useful sufficiently. Therefore, in order to validate the accuracy of our model, the following two types of verification experiments were conducted using incorrectly-predicted tone-mapping methods. Table.2 shows the results of the verification experiment 1 and 2.



**Fig.7 Example of prediction of incorrect TM method**

## 5.1 Verification Experiment 1 and Results

Verification experiment 1 employed 28 test images that failed to predict the tone-mapping method. The images of the predicted and the ground-truth tone-mapping method were displayed. The subjects compared the two images and answered with a three-grade evaluation described as follows.

- -1 : "The image of the ground-truth method is preferable"
- 0 : "The appearance of the two images is the same"
- 1 : "The image of the predicted method is preferable."

This experiment was conducted with 10 subjects. The score of each image was calculated based on the average and the standard error of 10 subjects. If the score range (average with standard error) falls within  $[-1, 0)$ , the tone-mapped image by the ground-truth method is more preferable. On the other hand, if the score range falls within  $(0, 1]$ , the tone-mapped image of the predicted method is more preferable. If the score range covered 0, both images have the same preference.

As a result of verification experiment 1, it was found that 11 out of 28 test images can be regarded as "similar to the appearance of both images" or "the image of the predicted tone-mapping method is more preferable." For the remaining 17 test images, our model can be regarded as failing to predict.

## 5.2 Verification Experiment 2 and Results

The verification experiment 2 was conducted using 17 images that resulted in "the image of the ground-truth tone-mapping method is more preferable." We examined preferable score of the tone-mapped image using our prediction model by comparing to the ground-truth method.

The subject compared the two images and responded to the tone-mapped image of our model with a five-grade evaluation described below. As in verification experiment 1, the image score range was calculated from the average score and standard error of 10 subjects.

- 5 : The two images have the same appearance and preference.
- 4 : The appearance between the both image are different, but the preference is almost same.
- 3 : The appearance between the both image are different, and the preference is slightly inferior.
- 2 : The appearance between the both image are different, and the preference is inferior.
- 1 : The appearance between the both image are different, and the preference is very inferior.

If the score range (average with standard error) is over 4.0 or covered 4.0, the preference of the tone-mapped image by our model is inferior to that of the ground-truth image, but the preference of our model is acceptable.

As shown in Table.2 of verification experiment 2, it was found that the scoring range of 7 out of 17 images fell within an acceptable range

Verification 1 shows the 11 images by our method provide more preferable appearance. Verification 2 shows

7 tone-mapped images by our method were acceptable compared to the ground-truth tone-mapped images. In Section 4.2, 17 images were predicted correctly. In summary, our system can properly predict the tone-mapping method with the accuracy of 78% (35 out of 45 images).

**Table.2 Results of verification 1 and 2**

Verification1 (28 images)		Verification2 (17 images)	
Grand truth is preferred.	17	Within acceptable limit.	7
Same	11	Within patience limits.	3
Predicted one is preferred.		Out of patience limit.	7

## 6 CONCLUSION AND FUTURE WORK

In this paper, we constructed an HDR image database by human evaluation experiments. We also implemented a system that could automatically select the best tone-mapping method by training a CNN model. Finally, we confirmed the accuracy through verification experiments. The results also showed that human subjective preference could be estimated with the accuracy of approximately 80%. As a result, it was suggested that the proposed system could select a tone-mapping method that provides a preferable appearance.

Our future work is to obtain more robust evaluation data to improve prediction accuracy. We would like to improve the prediction model for individual preferences by using a fine-tuning with individual evaluation data.

## REFERENCES

- [1] E.Reinhard, G.Ward, S.Pattanaik and P.Debevec, "HIGH DYNAMIC RANGE IMAGING", Morgan Kaufmann Publisher, pp.187-387, 2005.
- [2] P.Fersch, I.Tastl, and W.Purgathofer.A "comparison of techniques for the transformation of radiosity values to monitor colors." In First IEEE International Conference on Image Processing, pp.992-116, 1994.
- [3] E.Reinhard, M.Stark, P.Shirley, and J.Ferwerda, "Photographic tone reproduction for digital images", ACM Transactions on Graphics, 2002
- [4] F.Drago, K.Myszkowski, T.Annen, and N.Chiba, "Adaptive Logarithmic mapping for displaying high contrast scenes." Computer Graphics Forum, 2003.
- [5] M.Ashikhmin. "A tone mapping algorithm for high contrast images", In Proceedings of 13th Eurographics Workshop on Rendering, 2002.
- [6] H.Yee and S.Pattanaik, "Segmentation and adaptive assimilation for detail preserving display of high-dynamic range images." The Visual Computer, 2003.
- [7] K.Simonyan and A.Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition", ICLR 2015.