RESEARCH ON SINGLE IMAGE FUSION-BASED HIGH DYNAMIC RANGE ENHANCEMENT

Yanting Miao[§], Zhijie Wang[§], Kuanhan Shi, Yuling Hu

{y43miao, z2458wan, kuanhan.shi, y434hu}@uwaterloo.ca

ABSTRACT

High Dynamic Range (HDR) enhancement has recently become widely researched and used, especially the single image fusion-based HDR enhancement. Many fusion-based algorithms have been proposed while their applicability haven't been capture. In this paper, we reviewed six major fusion-based single image HDR enhancement methods, and evaluated and compared their enhanced results on two novel datasets. Hence, we pointed out the shortcomings of existing algorithms, and provided some recommendations for the future work in such area.

Index Terms— Single Image Enhancement, High Dynamic Range, Exposure Fusion

1. INTRODUCTION

Smart phones and digital cameras have been widely used to take pictures in our daily life. However, people often cannot get a satisfied photograph since the limitation of dynamic range of devices. In addition, in undesirable illuminated environments, digital cameras capture over-exposure or under-exposure images. For example, in highlight regions, those images may exhibit artifact halos Thus, in order to produce an image with better quality in undesirable illuminations, many image enhancement methods have been studied in last decades.

To improve high dynamic range in a single image, High Dynamic Range Imaging (HDRI) is a technique that has been used in imaging to achieve unusually high dynamic range and ameliorate quality of image. In the static scene, multiple image-based methods can generate HDR-effect images successfully, but, these methods cannot handle dynamic scene. If an objects was moving during the exposure time, multiple image-based approaches may produce ghost artifacts. One solution is to set up a multi-camera capturing system, whereas it is highly cost. For these reasons, a single image-based HDR algorithm is needed for some devices and scenarios. Recently, increasing researchers developed single image fusion-based HDR enhancement methods. The process of single image-based methods can be simply summarized as 1) generating multiple exposure pseudo-images; 2) selecting and adjusting exposure pseudo-images; 3) fusing pseudo-images to produce a HDR-effect image. Researchers have improved the method of generating exposure virtual images and developed fusion-algorithms for synthesizing images in past years. Nevertheless, several issues are not addressed in previous reviews. First, these existing fusion-based methods lacks experiments on novel and latest datasets. Second, these methods also miss clear and objective evaluation of the quality of produced images.

Therefore, to address these issues, in this project, our contributions to the study of single image fusion-based HDR enhancement are mainly about:

1. We reviewed 6 major single image fusion-based HDR enhancement algorithms. For the purpose of carrying out experimental-based comparisons, we re-implemented them on MATLAB.

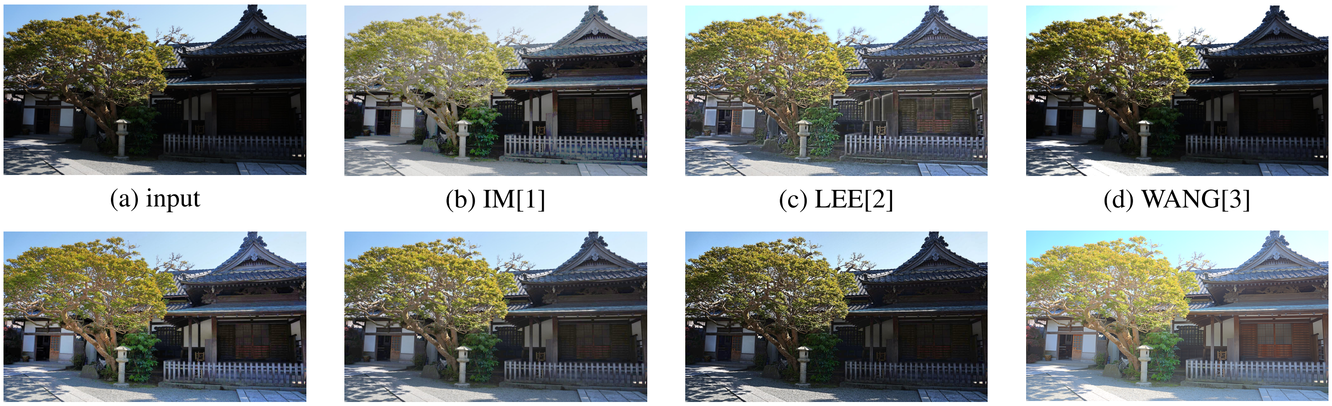
2. We carried out the comparisons on two latest datasets, which haven't been tested on these algorithms. Thus, we provided a critical evaluation of these methods based on enhanced images' quality and computing time, and shared our lessons-learnt and pointed out the shortcomings of these methods and problems faced in this area.

The rest of this paper is organized as following. Section 2 provides a review of single image fusion-based HDR enhancement methods. The experimental-based comparisons and critical discussions are provided in Section 3. And our conclusion and future work are proposed in Section 4

2. SINGLE IMAGE FUSION-BASED HDR ENHANCEMENT

There are numerous image enhancement techniques have been proposed to improve high dynamic range of degraded images. There are six major existing and state of the art algorithms that have similar fusion-based enhancement procedures. In 2011, Im *et al.*[1] proposed a single image-based ghost-free HDR method. This method stretches local histogram and edge-preserving spatially. One year later, Lee *et al.*[2] proposed a classified virtual exposure image fusion

[§]Equal contribution



(a) input (b) IM[1] (c) LEE[2] (d) WANG[3]

(e) FU[4] (e) YING[5] (f) PARK[6] (e) ground truth HDR

Fig. 1. The comparison on enhanced result from different algorithms on one image from HDR Eye dataset[7] and corresponding input and ground truth.

method that adjusts discrete wavelet transform (DWT) coefficients of images to produce a high-contrast image. However, Lee *et al.*[2]'s method suffer from its high time complexity of classifying each pixel, in this algorithm, each pixel has to be labeled, which is a time-consuming process. Later on, Wang *et al.*[3] proposed local region segmentation-based enhancement of exposure images to generate HDR image, which is similar to Lee *et al.*[2] classified method in some degree. But, in Wang *et al.*[3]'s approach, over-exposure scenes were also enhanced, and thus, this method cannot handle over-exposure images. To pursue better performance, Fu *et al.*[4] proposed to fuse the derived inputs with the corresponding weights in a multi-scale strategy to produce adjusted illumination. Recently, an image contrast enhancement algorithm that fuses input image and synthetic image according to the optimization-based weight matrix to obtain the enhancement result was proposed by Ying *et al.*[5]. Another algorithm was proposed by Park *et al.*[6] in 2019, which decomposes luminance into reflectance and illumination, and to apply selective reflectance scaling (SRS) method for the adjustment of reflectance. One example of enhanced image through these methods are shown in Fig 1.

3. EXPERIMENTS AND DISCUSSION

To study current research and development on fusion-based HDR enhancement, in this section we will firstly introduce the datasets and quality assessment metrics we used, and then compare the methods we have notified before, discuss about their advantages and disadvantages.

The compared methods of fusion-based HDR enhancement include: IM[1], LEE[2], WANG[3], FU[4], YING[5] and PARK[6].

In order to make our evaluation fair, all these methods are implemented on MATLAB R2018a, and all the experiments are running on a OS X 10.15.1 PC with 2.4 GHz Quad-Core Intel Core i5 CPU, 16 GB 2133 MHz LPDDR3 Memory.

3.1. Datasets

Most of previous works only evaluated their algorithms on several images, in [4] and [6], the authors only evaluated and compared their algorithms on about 10 images, which is too weak to show the performance on robustness of the algorithms. Besides, in [1] and [2], the authors only compared the enhanced images with un-enhanced images visually, which is not convincing enough. In [4] and [6], both of the authors used GMSD metric to compare the similarity between input image and enhanced image, however, since the dynamic range of images have been adjust and their luminance have been changed, it's not persuasive enough to evaluate the enhanced result on computing GMSD between input images and enhanced images.

Hence, in our experiments we use two novel datasets, HDR Eye dataset [7], and Source Identification HDR (SI-HDR) dataset [8], which haven't been tested on these methods at the same time. In [6], the authors evaluated their algorithms on a few images from HDR Eye dataset, but not the full one, and were not compared with other fusion-based methods we notified above. HDR Eye dataset was collected from camera, and presented both LDR image (.tiff format) and HDR (.hdr format) image from multiple-images-fusion. SI-HDR dataset was collected with several different mobile devices of major brands, and presented both LDR and HDR in .jpeg format. Since the input and enhanced images are in a standard dynamic range, we use MATLAB to tonemap the HDR images in HDR Eye dataset to a normal dynamic range one. All the images are pre-processed into 8bit JPEG file. In addition, based on different six mobile devices (iPhone 8, iPhone SE, iPhone 7, iPad Air, iPhone 6, and iPhone

5s), we split SI-HDR dataset into six categories (SI-HDR1, SI-HDR2, SI-HDR3, SI-HDR4, SI-HDR5, and SI-HDR6).

3.2. Image Quality Assessment Metric

It's difficult to find a general method to assess the quality of an image and the performance of enhancement algorithm. Therefore, we select a full reference IQA to compare enhanced result with ground truth HDR image, and two no reference IQAs to evaluate the quality of enhanced results.

For the full reference IQA, we select GMSD (gradient magnitude similarity deviation)[9] to evaluate the visual distortion between input and enhanced result, a lower GMSD value represent higher similarity between them. Fig.2 displays the average GMSD score of six different algorithms in our datasets. According to the average GMSD score of each algorithm corresponding to each dataset, we observe that YING[5] achieves the lowest score in 5 of 6 datasets, and in HDR Eye dataset, YING[5] wasn't the lowest one, but was very close to the lowest one with FU[4] (0.1500 vs 0.1438). For each dataset, the variance of each method's GMSD score were very close, which states that all these algorithms are robust.

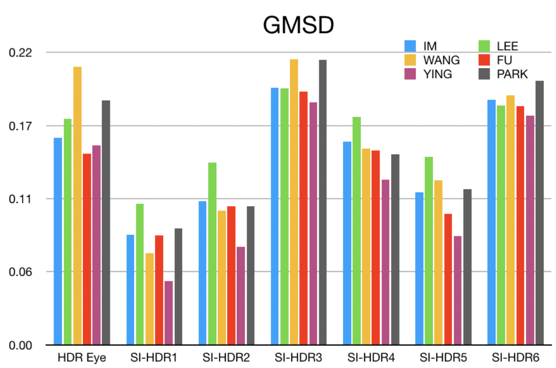


Fig. 2. Average GMSD score of each dataset & algorithm

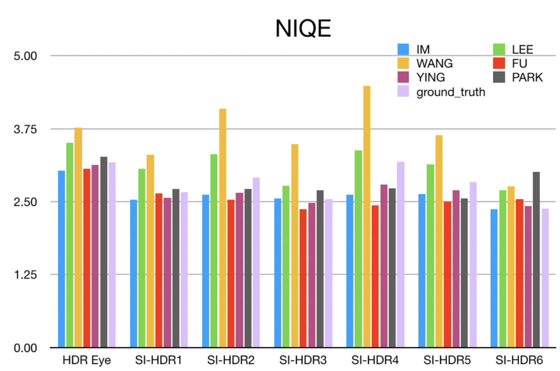


Fig. 3. Average NIQE score of each dataset & algorithm

For the no reference IQA, we select NIQE (Natural image quality evaluator)[10] and CORNIA (Codebook Representation for No-reference Image quality Assessment)[11]. NIQE is based on the constrution of natural scene statistic (NSS) model, which collected statstic features from natural and undistorted images. A smaller NIQE value implies the enhanced image has less distortions such as noise and blur. The average NIQE value of each algorithm in different datasets are shown in Fig.3. From Fig.3, we are clear that IM[1], FU[4], and YING[5]'s average NIQE score for each dataset are all close to the average NIQE score of ground truth HDR images. In our view, many images enhanced by WANG[3] have over-exposure areas, therefore, the average NIQE score of WANG[3] is much higher than others.

Different with other two metrics, in CORNIA metric, the score meaning depends on the database that we used. If the database use differential mean opinion score (DMOS) as labels of images, then a lower CORNIA score means the image has a better quality. However, if the database used mean opinion score (MOS) as labels of images, then a higher CORNIA score indicates a higher image quality. In our experiments, we use IQA LIVE Challenge [12] as the database of CORNIA, and this database adopts MOS as labels. The average CORNIA scores of these algorithms are shown in Fig.4. In the experiment of CORNIA, we also notice the images that generated by [1], [4], and [5] can have higher image quality than other approaches. Also, we observed that the variance of YING[5]'s CONIRA score in different datasets is much lower than others, which means that YING[5] is much more stable.

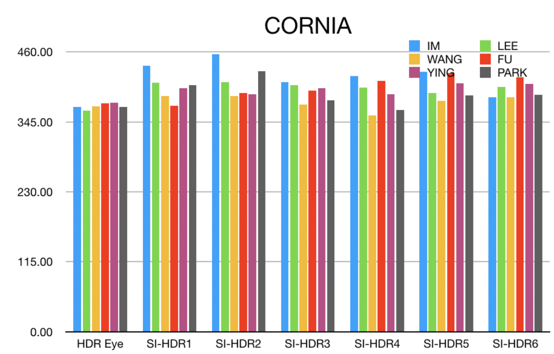


Fig. 4. Average CORNIA score of each dataset & algorithm

3.3. Run-time Comparison

Run-time comparison results are shown in Table 1 for 4 different resolutions in our datasets. With the increment of resolution, LEE[2] needs significantly more time than other 5 methods, which needs minutes instead of seconds. We believed it was caused by the computational cost on classifying the image pixels.

Image size	IM[1]	LEE[2]	WANG[3]	FU[4]	YING[5]	PARK[6]
684×484	0.06	16.64	0.21	0.17	0.11	0.66
816×612	0.09	26.64	0.34	0.26	0.17	1.11
1008×756	0.15	42.35	0.53	0.46	0.28	1.80
1920×1080	0.41	115.93	1.43	1.24	0.77	5.43

Table 1. Run-time (s) in seconds of each algorithm.

3.4. Results and Discussion

Combining the performance in GMSD, NIQE, CORNIA, and run-time testing, we think that the best three methods are algorithms of IM [1], YING [5], and FU [4]. The reason why [1] and [5] can achieve better results than other algorithms is mainly their process of generating pseudo-images. In Im *et al.*[1]'s algorithm, the input LDR image was transformed to a corresponding histogram. Im *et al.* splited this histogram to two subsets, which corresponds to a over-exposure image and a under-exposure image. In [1] based on input image and two pseduo-images, to generate a HDR-effect image. In Ying *et al.*[5]'s algorithm, the author optimized the exposure ratio, according to this optimized parameter to generate pseduo-images. Fu *et al.*[4]'s algorithm focus on fusion methods. In Fu *et al.* [4], the author proposed a "multi-scale fusion" method, which adopt Laplacian pyramid to extract features from input pseduo-images, and introduced Gaussian pyramid to smooth the weight function.

4. CONCLUSION

In this research, we review six existing HDR enhancement fusion-based algorithms and describe their characteristics. Then, we introduce two novel HDR datasets to compare these algorithms. Finally, we provide a clear evaluation of image quality of generated HDR-effect images via multiple image quality assessment metrics, including GMSD, NIQE, and CORNIA. In conclusion, we presents outline of HDR-effect image generation and provide experiment methods and comparison of fusion-based algorithms under two latest datasets, which help researchers to develop this area in the future. Furthermore, we believe the future work should focus on fusion method, and construct a sufficient HDR enhancement benchmark dataset to evaluate the perceptual quality of image.

5. REFERENCES

[1] Jaehyun Im, Jaehwan Jeon, Monson H Hayes, and Joonki Paik, "Single image-based ghost-free high dynamic range imaging using local histogram stretching and spatially-adaptive denoising," *IEEE Transactions on Consumer Electronics*, vol. 57, no. 4, pp. 1478–1484, 2011.

[2] Chang-Hsing Lee, Ling-Hwei Chen, and Wei-Kang Wang, "Image contrast enhancement using classified virtual exposure image fusion," *IEEE Transactions on Consumer Electronics*, vol. 58, no. 4, pp. 1253–1261, 2012.

[3] Tsun-Hsien Wang, Cheng-Wen Chiu, Wei-Chen Wu, Jen-Wen Wang, Chun-Yi Lin, Ching-Te Chiu, and Jing-Jia Liou, "Pseudo-multiple-exposure-based tone fusion with local region adjustment," *IEEE Transactions on Multimedia*, vol. 17, no. 4, pp. 470–484, 2015.

[4] Xueyang Fu, Delu Zeng, Yue Huang, Yinghao Liao, Xinghao Ding, and John Paisley, "A fusion-based enhancing method for weakly illuminated images," *Signal Processing*, vol. 129, pp. 82–96, 2016.

[5] Zhenqiang Ying, Ge Li, Yurui Ren, Ronggang Wang, and Wenmin Wang, "A new image contrast enhancement algorithm using exposure fusion framework," in *International Conference on Computer Analysis of Images and Patterns*. Springer, 2017, pp. 36–46.

[6] Jae Sung Park, Jae Woong Soh, and Nam Ik Cho, "Generation of high dynamic range illumination from a single image for the enhancement of undesirably illuminated images," *Multimedia Tools and Applications*, vol. 78, no. 14, pp. 20263–20283, 2019.

[7] Hiromi Nemoto, Pavel Korshunov, Philippe Hanhart, and Touradj Ebrahimi, "Visual attention in ldr and hdr images," in *9th International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM)*, 2015, number CONF.

[8] Omar Shaya, Pengpeng Yang, Rongrong Ni, Yao Zhao, and Alessandro Piva, "A new dataset for source identification of high dynamic range images," *Sensors*, vol. 18, no. 11, pp. 3801, 2018.

[9] Wufeng Xue, Lei Zhang, Xuanqin Mou, and Alan C Bovik, "Gradient magnitude similarity deviation: A highly efficient perceptual image quality index," *IEEE Transactions on Image Processing*, vol. 23, no. 2, pp. 684–695, 2013.

[10] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, 2012.

[11] Peng Ye, Jayant Kumar, Le Kang, and David Doermann, "Unsupervised feature learning framework for no-reference image quality assessment," in *2012 IEEE conference on computer vision and pattern recognition*. IEEE, 2012, pp. 1098–1105.

[12] D. Ghadiyaram and A. C. Bovik, "Massive online crowdsourced study of subjective and objective picture quality," *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 372–387, Jan 2016.