Single Image Contrast Enhancer Using Convolutional Neural Network and Bright Channel Prior

Prof.M.Mary Babitha M.E.,
Assistant professor,
Arunachala College of Engineering for Women, Manavilai.

Ashmi.R,
M.E.Final Year,Communication Systems,
Arunachala College of Engineering for Women, Manavilai.

Abstract— Due to the poor lighting condition and limited dynamic range of digital imaging devices, the recorded images are often under or over exposed with low contrast. Most of previous single image contrast enhancement (SICE) methods adjust the tone curve to correct the contrast of an input image. Those method fail in revealing image details because of the limited information in a single image. In proposed convolutional neural network (CNN) is used to train a SICE enhancer and bright channel prior to enhance image. With the constructed data set, a CNN can be easily trained as the SICE enhancer to improve the contrast of an under-/over-exposure image. Experimental results demonstrate the advantages of our method over existing SICE methods with a significant margin.

Index Terms— Single image contrast enhancement, multi-exposure image fusion, convolutional neural network, bright channel prior.

I. INTRODUCTION

The natural scene with good contrast, vivid color and rich details is an essential goal of digital photography. The acquired images are often under or over-exposed images because of poor lighting conditions and the limited dynamic range the imaging device. Resulting the low contrast and low Quality images will not only degenerate the performance of many computer vision and image analysis algorithm. Contrast enhancement is an important step to improve the quality of recorded images and make the image details more visible.

Traditional single image contrast enhancement (SICE) techniques include those histogram-based algorithm which increase the contrast of an image by redistributing the luminous intensity on histogram, and Retinex based algorithms, which enhance the reflectance and illumination components of the image separately. These methods, however, are difficult to reproduce a high-quality image due to the complex natural scenes and the limited information in a single low contrast image. Thanks to the development of imaging devices, we are able to capture a sequence of multi-exposure images in a short time to fulfil the dynamic range of a scene. With the sequence, multi-exposure image fusion (MEF) and stack-based high dynamic range (HDR) imaging (with a following tone mapping operator) methods, [15] can be applied to blend the multiple images with different exposures into a perceptually more appealing image.

Generally speaking, MEF and stack-based HDR methods will produce images with better visual quality than those SICE methods since more information is available in the multi-exposure sequence. However, the acquisition of multi-exposure image will complicate the imaging
process, and camera shake or moving objects will lead to unpleasant fusion artifacts such as the the MEF method could recover more image details, which cannot be revealed by the SICE method; however, it generates some ghosting artifacts due to the displacement of different frames caused by the object motion (such as human movement and ripples). In contrast, the SICE method will not have such ghosting artifacts because it takes only one single exposure image as input. Due to the above reasons, SICE is more attractive and easier to implement in practice, yet it is much more challenging because of the limited information in a single image.

II. RELATED WORK

Single Image Contrast Enhancement

Single image contrast enhancement (SICE) aims to improve the visibility of the scene in a given single low-contrast image. It provides a way to enhance the low contrast photographs captured from a high dynamic range scene. Many histogram and Retinex based SICE methods have been proposed in the past decades. Histogram-based methods have been widely used because of their simplicity in enhancing low-contrast images. Those methods attempt to redistribute the luminous intensity on histogram in a global or local manner. However, such simple redistribution operations may produce serious unrealistic effects in the enhanced images since they ignore image structural information. To excavate the structural information from the low-contrast image, Retinex-based methods decompose the input image into albedo and illumination layers, and adopt different strategies to enhance the reflectance and illumination components. Most of the previous SICE methods are based on some assumptions on high-quality images, while they may not fully exploit the information in the input image. Since extra information can be learned from the external dataset, in this work, we will elaborately build a dataset to learn a powerful CNN-based SICE enhancer from multi-exposure images.

MEF and Stack-Based HDR

Because of the limited dynamic range, traditional digital imaging systems may lose structural details when shooting a natural scene. To address this issue, stack-based HDR methods propose to merge bracketed multiple exposure images into an HDR irradiance map, then employ a tone mapping operator to compress the dynamic range of HDR irradiance map so that the high-contrast image can be displayed on regular monitors. Different from the HDR approaches, MEF methods attempt to fuse the images directly in the non-linear brightness domain to reproduce a high-visibility image. Despite their successes on well-aligned image sequences, the existence of camera shake and object motion in many scenes often leads to ghosting artifacts in the final enhanced results, limiting the applications of MEF and HDR in practice. In the last decade, researchers have spent much efforts to design de-ghosting algorithms and learning-based methods like proposed to map the multi-exposure image sequences to an HDR image. However, it remains a challenging problem in MEF and HDR for dynamic scenes. In this work, we elaborately select the well-aligned image sequences to generate a good reference images by MEF and HDR reconstruction methods.

CNN for Image Restoration

CNN has demonstrated its effectiveness in image restoration and enhancement applications such as denoising [34], [35], super-resolution [21] and deblurring [36]. In those applications, pairs of degraded images and their high-quality counter parts can be easily generated. With those paired training data, CNN can be used to learn a mapping function between the degraded observations and their corresponding high-quality reference images. However, for the application of SICE, such rule-based, computer-generated training datasets are too ideal to be true for real-world low-contrast images, where the distribution of luminance is much more complex and varies with different scenes, cameras and camera settings.
III. MULTI-EXPOSURE IMAGE DATASET AND REFERENCE IMAGE GENERATION

Objectives

There are two major objectives for our multi-exposure image dataset construction. First, the dataset should contain enough high resolution multi-exposure image sequences and cover a diversity of scenes. Second, for each sequence, a high-quality reference image should be generated so that image pairs can be constructed for end-to-end learning.

Multi-Exposure Image Collection

To achieve the objectives mentioned above, we collect and select multi-exposure image sequences of relatively static scenes. The details of data collection and screening are described as follows.

1) Data Collection: To ensure that a robust and general SICE enhancer can be trained, the training data should be collected from representative real-world scenarios with commonly used imaging devices. Some multi-exposure image sequences are available in literature [37], [38] and most of them taken by different cameras and from different were captured for the study of MEF and stack-scenes. Seven types of consumer grade cameras are based HDR methods. However, the total amount of used to collect the image sequences, including Sony such publicly available sequences is very limited, α7RII, Sony NEX-5N, Canon EOS-5D Mark II, and many of them were taken under indoor environment, Canon EOS-750D, Nikon D810, Nikon D7100 and many. Neither the number of sequences nor the iPhone 6s. diversity of sequence exposure levels meets the requirements of real-world applications. To achieve the first objective, we collect a large number of sequences from both indoor and outdoor scenes, and make sure that the photographs in our dataset cover a broad range of scenes, subjects and lighting conditions. Some sample sequences of our multi-exposure image dataset are presented in Figure 1. Some sequence are from, while the others are collected.

Some multi-exposure image sequences are devices. In our dataset, the image sequences are available in literature [37], [38] and most of them taken by different cameras and from different were captured for the study of MEF and stack-scenes. Seven types of consumer grade cameras are based HDR methods. However, the total amount of used to collect the image sequences, including Sony such publicly available sequences is very limited, α7RII, Sony NEX-5N, Canon EOS-5D Mark II, and many of them were taken under indoor environment, Canon EOS-750D, Nikon D810, Nikon D7100 and many. Neither the number of sequences nor the iPhone 6s. diversity of sequence exposure levels meets the requirements of real-world applications. To achieve the first objective, we collect a large number of sequences from both indoor and outdoor scenes, and make sure that the photographs in our dataset cover a broad range of scenes, subjects and lighting conditions. Some sample sequences of our multi-exposure image dataset are presented in Figure 1. Some sequence are from, while the others are collected.
see in our daily life. In this work, we collect image sequences from both indoor and outdoor scenes under certain EV settings. For the indoor scenes, we are able to set up a static environment and use a tripod to capture well-aligned image sequences. We collect 7 to 18 images for each indoor scene. The exposure levels are manually set based on the lighting ratio of the scene. For the uncontrolled outdoor environment, moving objects (e.g., cars, walking people, shaking trees) make the acquisition of well-aligned sequences very challenging. We use the continuous bracket mode to automatically shoot image sequences with shifted exposures. After collecting the source images, a further screening process is conducted to select desirable sequences for reference image generation.

Data Screening: In the data collection stage, we collected more than 10,000 image sequences with different exposure levels (including repeated sequences). However, many of them contain distorted images or moving objects. We therefore conduct an uphill screening process to refine the dataset. Sequences with significantly distorted images (e.g., motion blur, out of focus and visible sensor noise) or obvious moving objects are discarded (see Figure 2 for some examples).

![Fig. 2. Sample source image sequences excluded from the dataset. Blue and red arrows point out the moving objects in different frames, which will cause ghosting artifacts in the reference image generated by MEF or stack based HDR algorithms.](image)

Having the candidate sequences, we propose to generate high-quality reference images with MEF and stack-based HDR methods. 13 state-of-the-art MEF and HDR algorithms are employed in this process including 8 MEF methods. The implementations of those algorithms are obtained from the original authors, except for Raman09 and Bruce14 whose implementations are from an HDR toolbox in Github [46]. To generate the faithful results of original scenes and for a fair comparison, we manually tune the tone-mapping operators for Sen12, Hu13 and Oh15 by Photomatix. As a result, there is one image for each scene by each method for subjective evaluation. Note that we also use Photomatix to generate the HDR irradiance map for Hu13, which outputs a stack of well-aligned low dynamic range images. As for Raman09 and Bruce14, we adopt the tone-mapping operators as presented in the original papers.

Reference image generation

![Fig. 3. User interface for subjective testing](image)

IV. CNN-BASED SICE LEARNING

With the constructed dataset, we can design a CNN based SICE enhancer to learn a mapping function $I(x, y)$ between the low-contrast input image $I(x, y)$ and its corresponding reference image. One can directly train a deep CNN $H(I, W)$ with parameters $W$ to achieve this goal. We train the direct network with mean square error loss, l-norm loss and structural dissimilarity loss respectively. The MSE loss function to be minimized is

$$l_2(w) = \frac{1}{n} \sum_{i=1}^{n} \| I_i^{(t)} - H(I_i^{(0)}, W) \|^2_f.$$
A. Network Overview

The proposed CNN has 5 types of layers which are shown in Figure 8 with 5 different colors. i) Conv+PReLU: 64 filters of size $3 \times 3$, $5 \times 5$ and $9 \times 9$ with strides 1 and 2 are used to generate 64 feature maps, and PReLU (parametric rectified linear unit) [55] is utilized for the nonlinearity. ii) Deconv+PReLU: 64 filters of size $9 \times 9$, $5 \times 5$ and $3 \times 3$ with strides 2 and 1 are used to generate 64 feature maps, and PReLU is utilized as the activation function. iii) Conv+BN+PReLU: 64 filters of size $3 \times 3$ are used, and batch normalization [56] is added between convolution and PReLU. iv) Conv: 3 filters of size $1 \times 1$ are used to reconstruct the output. v) Skip connection: the add operation is used to connect the feature maps of two layers.

1) Stride Convolution and Deconvolution: The convolutional operations will reduce the size of feature maps. To ensure that the output image will have the same size as the input one, methods have been proposed to pad zeros before convolutions [21]. However, for the luminance enhancement network, we experimentally found that padding zeros would lead to artifacts around the boundary of the output image. Therefore, instead of padding zeros, we apply the deconvolutions to keep the size of the output unchanged. The convolutional and deconvolutional strategy not only avoid artifacts in the boundary area, but also reduce computational burden with stride filters.

2) Parametric Rectified Linear Unit: In many CNN-based image restoration methods [21], [23], rectified linear unit (ReLU) is adopted as the activation function. However, since both the positive and negative coefficients contain important local structural information of the input image, simply setting the negative responses to zeros may not be a good choice. In this paper, we adopt the PReLU as the activation function, which could improve model fitting with nearly zero extra computational cost and little over-fitting risk [55]. Without ignoring negative coefficients, PReLU is able to generate high quality estimation with less filters. In Figure 9, we provide some visual examples to compare the two activation functions.

B. Component Enhancement Network

1) Luminance Enhancement Network: In order to increase the contrast of luminance map $L$ and enforce spatial smoothness on it, we train a luminance enhancer to learn a mapping function between the luminance component $L_{\text{original}}$ of the input low-contrast image and the luminance component of the reference image. Since the luminance component represents the global naturalness, the local receptive fields of the network is larger to connect with more pixel in the original
image. In order to increase the receptive fields while preventing the loss of information caused by the stride convolution operation, we adopt U-net [57] as the luminance enhancement network.

2) **Detail Enhancement Network**: A CNN mapping function $F_R$ with parameters is trained enhance the detail component $R_{\text{original}}$ of the input low contrast image to detail component of the reference image. Inspire by adopted residual networks. According to this structure guarantees that the input informations can be propagate through all parameter layer which helps the train network. The details of the proposed detail enhancement network are summarized in the Table II

---

**C. Whole Image Enhancement Network**

By using luminance and detail enhancement CNNs, able to enhanced the luminance ranges and also may recover more detail of the original low contrast image. However we cannot ensure overall visual quality of the enhance image because the two CNNs are trained separately on luminance components.

---

**TABLE I**

**LUMINANCE ENHANCEMENT NETWORK ARCHITECTURE**

| Layer       | Activation size |
|-------------|-----------------|
| Input       | $129 \times 129 \times 3$ |
| $3 \times 3 \times 64$ conv, stride 1, pad 1 | $129 \times 129 \times 64$ |
| $3 \times 3 \times 64$ conv, stride 1, pad 1 | $129 \times 129 \times 64$ |
| $3 \times 3 \times 64$ conv, stride 1, pad 1 | $129 \times 129 \times 64$ |
| $3 \times 3 \times 64$ conv, stride 1, pad 1 | $129 \times 129 \times 64$ |
| $3 \times 3 \times 64$ conv, stride 1, pad 1 | $129 \times 129 \times 64$ |
| $3 \times 3 \times 64$ conv, stride 1, pad 1 | $129 \times 129 \times 64$ |
| $3 \times 3 \times 64$ conv, stride 1, pad 1 | $129 \times 129 \times 64$ |
| $2 \times 1 \times 3$ conv, stride 1 | $129 \times 129 \times 3$ |

**Residual sum**

| Layer       | Activation size |
|-------------|-----------------|
| $129 \times 129 \times 3$ |

---

**TABLE II**

**DETAIL ENHANCEMENT NETWORK ARCHITECTURE**

| Layer       | Activation size |
|-------------|-----------------|
| Input       | $129 \times 129 \times 3$ |
| $3 \times 3 \times 64$ conv, stride 1, pad 1 | $129 \times 129 \times 64$ |
| $3 \times 3 \times 64$ conv, stride 1, pad 1 | $129 \times 129 \times 64$ |
| $3 \times 3 \times 64$ conv, stride 1, pad 1 | $129 \times 129 \times 64$ |
| $3 \times 3 \times 64$ conv, stride 1, pad 1 | $129 \times 129 \times 64$ |
| $3 \times 3 \times 64$ conv, stride 1, pad 1 | $129 \times 129 \times 64$ |
| $3 \times 3 \times 64$ conv, stride 1, pad 1 | $129 \times 129 \times 64$ |
| $2 \times 1 \times 3$ conv, stride 1 | $129 \times 129 \times 3$ |

**Residual sum**

| Layer       | Activation size |
|-------------|-----------------|
| $129 \times 129 \times 3$ |
V) EXPERIMENTAL RESULTS

We evaluate the proposed CNN enhancer on the built multi-exposure image dataset as well as images outside the dataset. We first present the experimental settings, and then present the comparison results with state-of-the-art SICE methods, as well as MEF and stack-based HDR methods.

A. Experimental Setting

We split all the 589 sequences randomly into training, validation, and test sets with a ratio of 7:1:2. All the three sets are guaranteed to contain images from indoor and outdoor scenes, which contain images with different exposure levels. Note that to further demonstrate the robustness of our method, we also conduct experiments on images outside our dataset, specifically, images from [14]. We use 720, 128 patches of size 129×129 are cropped from the training images, and stochastic gradient descent (SGD) with a batch size of 80 patches is used in training. We implement our model using the Tensor-Flow package. The momentum parameter and weight decay parameter are set to 0.9 and 0.0001, respectively. The method described in [55] is employed to initialize the weights.

VI. SIMULATION RESULTS

The proposed method uses multi-exposure dataset construction in the input image. The edge extraction process can be done by edge detection method. The Multi-Exposure Image Fusion (MEF) and stack-based High Dynamic Range (HDR) methods are used. The high quality reference images of input image and their corresponding histograms are shown in fig 8. The CNN based SICE enhancer is used to fuse the output image. The input low contrast and the output high contrast images are shown in fig 10.

Fig 6. Input image

Fig 7. Edge Extraction

Fig 8. Reference Image
VI. CONCLUSION AND FUTURE WORK

We built a multi-exposure image dataset, which has image sequences and 4,413 high-resolution images of different exposures. For each sequence, a corresponding high quality reference image was generated by using 13 MEF and stack-based HDR algorithms. Subjective tests are also conducted to screen the best quality one as the reference image of each scene. The availability of low-contrast images and their high-quality reference images in our dataset allows the end-to-end learning of high performance SICE methods. As a demonstration, we developed a simple yet powerful CNN-based SICE enhancer, which is capable of adaptively generating high quality enhancement result for a single over-exposed or under-exposed input image. Our experimental results showed that the developed SICE enhancer significantly outperforms state-of-the-art SICE methods, and even outperforms MEF and stack-based HDR methods for dynamic scenes.

REFERENCES

[1] X. Guo, Y. Li, and H. Ling, “LIME: Low-light image enhancement via illumination map estimation,” IEEE Trans. Image Process., vol. 26, no. 2, pp. 982–993, Feb. 2017.

[2] X. Fu, D. Zeng, Y. Huang, X.-P. Zhang, and X. Ding, “A weighted variational model for simultaneous reflectance and illumination estimation,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2016.

[3] T. Arici, S. Dikbas, and Y. Altunbasak, “A histogram modification framework and its application for image contrast enhancement,” IEEE Trans. Image Process., vol. 18, no. 9, pp. 1921–1935, Sep. 2009.

[4] T. Celik and T. Tjahjadi, “Contextual and variational contrast enhancement,” IEEE Trans. Image Process., vol. 20, no. 12, pp. 3431–3441, Dec. 2011.

[5] S.-C. Huang, F.-C. Cheng, and Y.-S. Chiu, “Efficient contrast enhancement using adaptive gamma correction with weighting distribution,” IEEE Trans. Image Process., vol. 22, no. 3, pp. 1032–1041, Mar. 2013.

[6] D. J. Jobson, Z.-U. Rahman, and G. A. Woodell, “A multiscale Retinex for bridging the gap between color images and the human observation of scenes,” IEEE Trans. Image Process., vol. 6, no. 7, pp. 965–976, Jul. 1997.

[7] S. Wang, J. Zheng, H.-M. Hu, and B. Li, “Naturalness preserved enhancement algorithm for non-uniform illumination images,” IEEE
Trans. Image Process., vol. 22, no. 9, pp. 3538–3548, Sep. 2013.

[8] D. J. Jobson, Z.-U. Rahman, and G. A. Woodell, “Properties and performance of a center/surround Retinex,” IEEE Trans. Image Process., vol. 6, no. 3, pp. 451–462, Mar. 1997.

[9] P. E. Debevec and J. Malik, “Recovering high dynamic range radiance maps from photographs,” in Proc. ACM SIGGRAPH Classes, 2008.

[10] S. K. Nayar and V. Branzoi, “Adaptive dynamic range imaging: Optical control of pixel exposures over space and time,” in Proc. ICCV, 2003.

[11] S. Li, X. Kang, and J. Hu, “Image fusion with guided filtering,” IEEE Trans. Image Process., vol. 22, no. 7, pp. 2864–2875, Jul. 2013.

[12] T. Mertens, J. Kautz, and F. Van Reeth, “Exposure fusion: A simple and practical alternative to high dynamic range photography,” Comput. Graph. Forum, vol. 28, no. 1, pp. 161–171, 2009.

[13] P. Sen, N. K. Kalantari, M. Yacoubi, S. Darabi, D. B. Goldman, and E. Shechtman, “Robust patch-based hdr reconstruction of dynamic scenes,” ACM Trans. Graph., vol. 31, no. 6, p. 203, 2012.

[14] T. H. Oh, J. Y. Lee, Y. W. Tai, and I. S. Kweon, “Robust high dynamic range imaging by rank minimization,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 37, no. 6, pp. 1219–1232, Jun. 2015.

[15] K. Ma, K. Zeng, and Z. Wang, “Perceptual quality assessment for multi-exposure image fusion,” IEEE Trans. Image Process., vol. 24, no. 11, pp. 3345–3356, Nov. 2015.

[16] K. Ma, H. Yeganeh, K. Zeng, and Z. Wang, “High dynamic range image compression by optimizing tone mapped image quality index,” IEEE Trans. Image Process., vol. 24, no. 10, pp. 3086–3097, Oct. 2015.

[17] W. T. Freeman, E. C. Pasztor, and O. T. Carmichael, “Learning low-level vision,” Int. J. Comput. Vis., vol. 40, no. 1, pp. 25–47, 2000.

[18] V. Bychkovsky, S. Paris, E. Chan, and F. Durand, “Learning photo-graphic global tonal adjustment with a database of input/output image pairs,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2011, pp. 97–104.

[19] S. B. Kang, A. Kapoor, and D. Lischinski, “Personalization of image enhancement,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2010, pp. 1799–1806.

[20] J. Kim, J. Kwon Lee, and K. Mu Lee, “Accurate image super-resolution using very deep convolutional networks,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2016, pp. 1646–1654.

[21] L. Xu, J. Ren, Q. Yan, R. Liao, and J. Jia, “Deep edge-aware filters,” in Proc. 32nd Int. Conf. Mach. Learn. (ICML), 2015, pp. 1669–1678.

[22] C. Dong, C. C. Loy, K. He, and X. Tang, “Learning a deep convolutional network for image super-resolution,” in Proc. Eur. Conf. Comput. Vis., 2014, pp. 184–199.

[23] V. Jain and S. Seung, “Natural image denoising with convolutional networks,” in Proc. Adv. Neural Inf. Process. Syst., 2009, pp. 769–776.
