AIS3/VHF5-1
Invited

Smart Display Clearly Sees via Deep Learning

Sehoon Lim[1], Luming Liang[1], Neil Emerton[1], Tim Large[1], Steven Bathiche[1]

sehoon.lim@microsoft.com
[1]Microsoft Applied Sciences, Redmond WA
Keywords: under-display camera (UDC); organic light-emitting diode (OLED); signal-to-noise ratio (SNR); modulation transfer function (MTF); deep neural network (DNN).

ABSTRACT

Under-display camera is of great interest in the display industry potentially eliminating the display bezel and camera notch/hole in mobile devices. However, display pixels cause complex signal modulation in the camera aperture which results in light diffraction and signal loss. We explored learning-based image restoration approaches to achieve high contrast and high modulation-transfer-function in modern image processing.

1 INTRODUCTION

World first under-display camera (UDC) phone was released in market by ZTE Axon 20 5G [1]. The selfie camera was implemented with a customized low-resolution and transparent display region centered on the camera location. Placing a camera under a display conflicts with the needs of high-quality camera imaging which conventionally requires a clear aperture to receive enough uninterrupted light from the scene. The display panel in front of the camera prevents fulfillment of the imaging requirements by modulating the incident light due to the display's 3D structure.

The display panel is typically a combination of stacked optical layers such as polarizers, pixel structure, and a substrate. The pixel structure consists of anode, various bandgap management layers like hole injection layer, OLED, and cathode. The anode is totally opaque (highly reflective): it is typically made of silver and indium tin oxide (ITO), and the shape of the anode defines the emitting area. The cathode is a thin metal layer that covers the whole area and is partially reflective.

BOE [2] is actively developing displays that are compatible with the UDC and we have evaluated a sample device in Figure 1. In BOE's sample the TFTs are not in the optical area, so the TFTs do not affect the UDC performance. It is the anode that primarily defines the diffraction in these kinds of UDC designs. Image degradation mainly arises from contrast reduction resulting from diffraction from the pixel pattern, and signal attenuation from the stacked layers in the display panel.

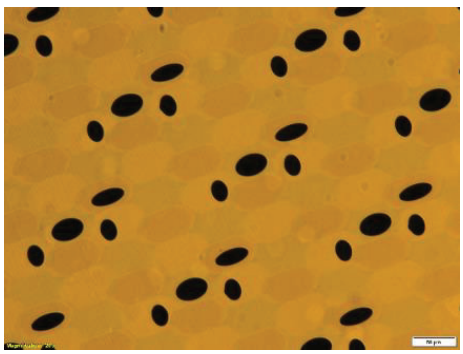

Fig. 1 UDC display sample

In Display Week 2020, we reported the introductory discussion that covers the full scope of problem definition, sample characterization, and learning-based image restoration [3]. In this paper, we generalize the DNN-based solution from the previous case study and suggest a more sophisticated approach used in conjunction with BOE's UDC design. While our previous solution was in research exploration, current idea could be more economic and feasible in practice. The learning-based method is updated by an advanced model that compactly recovers the corrupted camera image pixels from relevant neighboring pixels. The results are evaluated by the image metrics of modulation transfer function (MTF) and the signal-to-noise ratio (SNR).

2 IMAGE RESTORATION

Computationally, we de-link the denoising and deblurring processes for practical reasons. The main reason for wanting to split the deconvolution from the DNN is to de-couple the design of the pixels from the DNN training process, which can be time consuming and expensive. We do not want a heavy retraining of DNN model for different display panels. Also, UDC strongly suffers from photon insufficiency due to the light obscuration in case of low light conditions. UDC's blur kernel mainly depends on the display panel design and the blur kernel could be estimated using the known panel pattern. Since we know the blur degree in the UDC system, we separately take deblurring process that mitigates the computational workload in DNN.

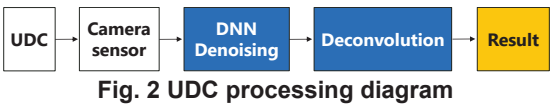

Fig. 2 UDC processing diagram

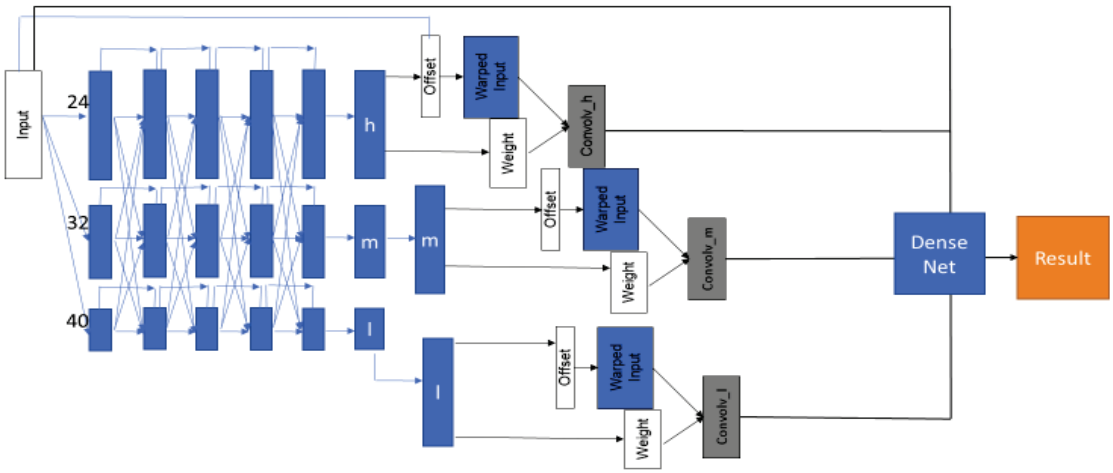

Fig. 3 DNN denoising structure

The DNN method first denoises the raw images from UDC and turns the deblurring process to deconvolution block as shown in Figure 2. In DNN denoising, HRnet-based deformable convolution was adopted to deal with severe noise impact in the UDC image. The HRnet repeatedly uses multi-scale fusion in high, medium, and low resolutions to recover corrupted pixels inferring from the different scales [4]. While Unet loses pixel-wise information in encoding and decoding paths [5], the HRnet keeps the high-resolution image along the network which avoids the pixel-wise information loss. The HRnet uses only three scales and the network's complexity is much smaller than the Unet. Right after the HRnet, Deformable convolution adaptively finds the most relevant neighboring pixels in noisy condition instead of using regular grid [6]. Every pixel is effectively learned by the most related pixels in neighboring content. At the end, the high, medium, and low-resolution features are merged to result in high-resolution image with denoising. The proposed model takes a 4-channel raw input and returns a 4-channel raw output in Figure 3.

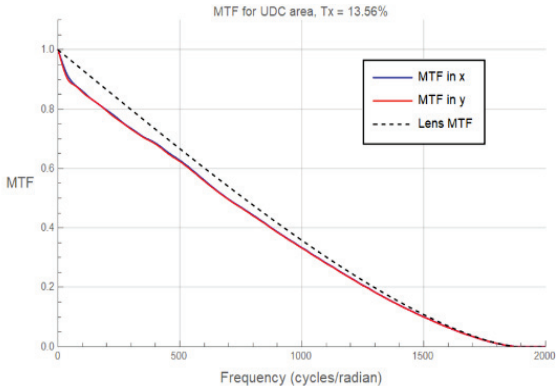

Fig. 4 UDC display characteristics of MTF

We collect 100 bursts of images of static scenes with a high-resolution still camera and each burst contains 100 images. We then take each image in the burst as a source image and the average image of each burst as the target image, forming a training dataset. To facilitate training, we crop both the source and target images into 256 x 256 patches, since the size of the receptive field is smaller than that. We trained the model using Adam optimizer [7], with learning rate = 0.01 x $0.5^{(\text{num of iterations})/10000}$. We set the batch size equals to 16 during the training.

The Wiener deconvolution method was used for deblurring. With this method the approximate spectral noise characteristics of the image are considered and

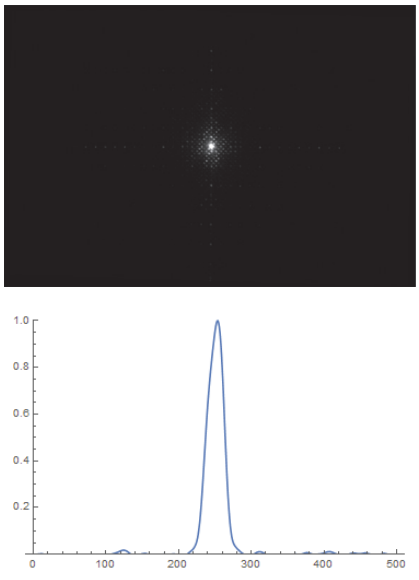

Fig. 5 Diffraction pattern (top) and line plot (bottom)

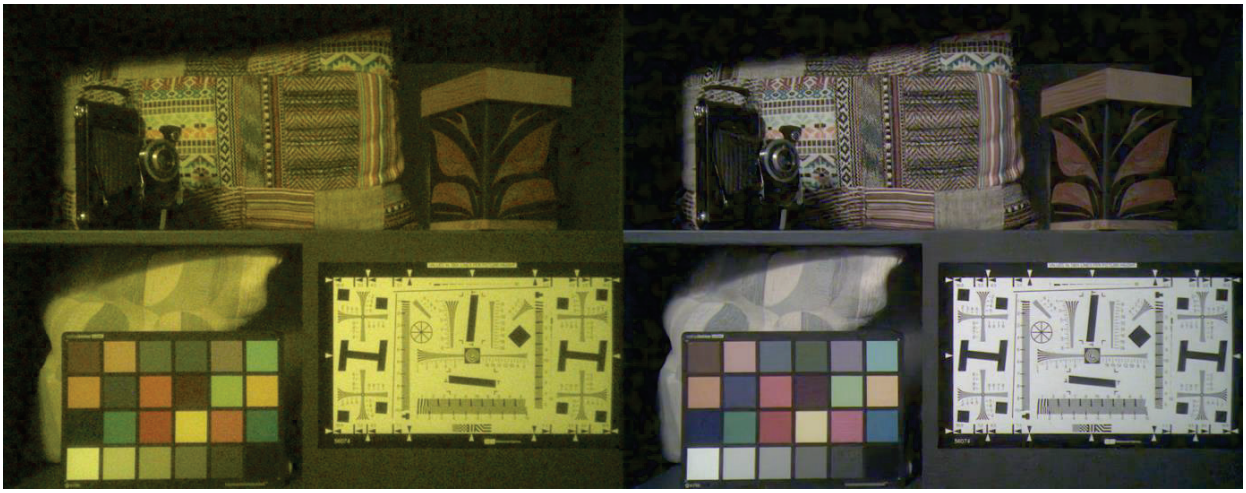

Fig. 6 Image restoration at 66 lux: through-screen corrupted (left-hand), and recovered (right-hand)

the filter weights the degree of denoising according to the SNR ratio at a given spatial frequency. Noisier spatial frequencies are therefore amplified less than frequencies with lower noise. This approach effects a compromise between increasing the resolution and increasing the overall noise.

3 EXPERIMENT

We tested a UDC sample from BOE which is consisted of an UDC area with high transmission and a normal display area with high pixel density. The microscope image already showed the UDC area in Figure 1 and the dark pattern with opacity floats in the background of OLED materials. In order to achieve about 20% transmission in the camera area, the pixel density was decreased by factor of 2, resulting in 200 PPI, and transparent tracks were used to carry current to the anode. The cathode is still the same (partially reflective everywhere over the whole display). The spectral transmission of UDC area varies in the visible range: the OLED display absorbs more light in blue than in red. The transmission characteristics indicate the impact of UDC could be severe depending on the spectral content and lighting condition of scene. The MTF of UDC sample was also measured in experiment (Figure 4). The UDC's MTF is slightly worse than the lens-only MTF and the degradation is less than 10% at any frequency. Figure 5 shows a diffraction pattern through the UDC sample using a Gaussian-filtered laser beam. The line plot of cross-section confirms most of diffraction energy falls into the zeroth order and higher orders are relatively low.

Experimental images are compared for through-screen corrupted and recovered images in Figure 6. A camera module in house was used with 1 micron pixel pitch and f-number 2. The through-screen corrupted image (left-hand) shows the blur due to diffraction and the noise due to low transmission from the display panel. The corrupted image is slightly yellow because of the spectral response of transmission. The recovered image (right-hand) was denoised by DNN and deblurred by Wiener deconvolution sequentially. The noise was effectively suppressed without sacrificing the high frequency features. For example, the uniformity of flat features was found in ColorChecker. At the same time, the high frequency features were also preserved in the texture of cushions. Both high SNR and high sharpness were achieved by de-coupling the denoising and the deblurring.

Figure 7 shows quantitative analysis for the SNR (top)

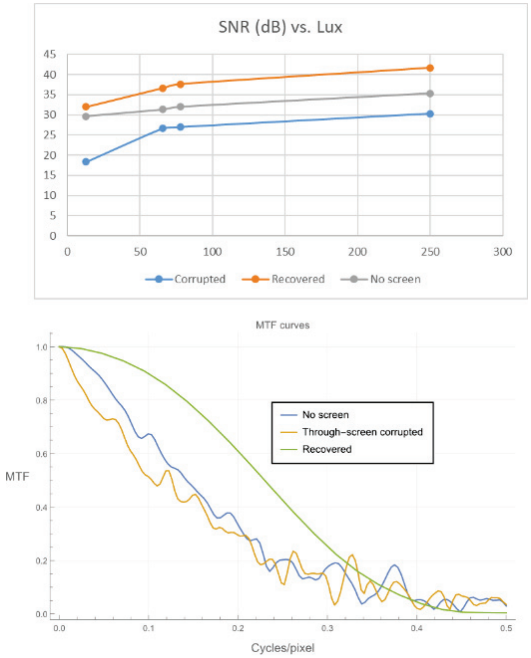

Fig. 7 Performance analysis: SNR (top), and MTF (bottom)

and MTF (bottom). The SNR was calculated by pixels of white color region in ColorChecker. The SNR of recovered image was improved by at least 10 dB compared to that of corrupted image. The SNR of recovered image was even higher than that of no screen image by up to 7 dB. To evaluate the image sharpness, we used the slant edge method that measures the spatial frequency response as an approximation of the MTF [8, 9, 10]. In the test, an ISO 12233:2000 resolution chart with slant edge features was captured through the display sample and the raw image was recovered by our image restoration method. In the figure, the MTF of recovered image was superior to those of corrupted image and no screen image. The MTFs of corrupted image and no screen image are noisy because they are not denoised at all.

4 CONCLUSIONS

We demonstrated high-resolution image restoration through a BOE's UDC sample using modern image processing methods. The performance of our method was analyzed by the standard camera metrics of MTF and SNR in sampled cases.

We still have corner cases for UDC performance in the display and camera, respectively, compared to conventional display and camera. Currently the pixel density of UDC is lower than the normal display and the camera suffers from flare and photon deficiency in high dynamic range. We continue to find the best solution for the two-fold problem that fundamentally trades off the display and camera experience.

5 ACKNOWLEDGEMENTS

The authors are grateful to BOE for providing a sample of UDC device and for helpful discussions on this project.

REFERENCES

[1] ZTE Axon 20 5G, https://www.zte.com.cn/global/about/news/2020081 7e1.html

[2] BOE, https://www.boe.com/cn/en/USA.html

[3] S. Lim, Y. Zhou, N. Emerton, T. Large, and S. Bathiche, "Image Restoration for Display-Integrated Camera," SID, Display Week, 74-1, (2020).

[4] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," CVPR, arXiv, 1505.04597 (2015).

[5] J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, and Y. Wei, "Deformable Convolutional Networks," CVPR, arXiv, 1703.06211 (2017).

[6] K. Sun, B. Xiao, D. Liu, and J. Wang, "Deep High-Resolution Representation Learning for Human Pose Estimation," CVPR, arXiv, 1902.09212 (2019).

[7] D. P. Kingma, and J. Ba, "Adam: A Method for Stochastic Optimization," CVPR, arXiv, 1412.6980 (2017).

[8] CPIQ P, "Standard for camera phone image quality," Institute of Electrical and Electronics Engineers (IEEE), (2015).

[9] ISO/TC42/WG18, "Resolution and spatial frequency response," International Organization for Standardization (ISO), (2000).

[10] ISO/TC42/WG18, "Resolution and spatial frequency response," International Organization for Standardization (ISO), (2014).