

A Realistic Dataset for Image Denoising and Demosaicking

May 14, 2017

Abstract

In this report, we introduce the motivation, idea, procedure, and results of our real noisy image dataset with “ground truth”. In this paper, we propose to simultaneously employing the spatial and temporal information to construct a real-world dataset which contain the original raw images as well as corresponding “ground truth” images. This dataset can be used in real world image denoising, demosaicking, and joint denoising and demosaicking (JDD) problems.

1 Introduction

According to [1], the noise corrupted in the imaging process is signal dependent and comes from five main sources: photon shot, fixed pattern, dark current, readout, and quantization noise. The noise model could be approximately modeled as

$$P = f((g_{cv}(\mathbf{C} + \mathbf{D}) + \mathbf{N}_{reset})g_{out} + \mathbf{N}_{out}) + \mathbf{Q} \quad (1)$$

Nam et al. [2] constructed a realistic dataset including real noisy images and corresponding “ground truth”. This dataset includes noisy images of 11 static scenes. The noisy images were collected under controlled indoor environment. Each scene was shot 500 times under the same camera and camera setting. The mean image of the 500 shots is roughly taken as the “ground truth”. Khashabi et al. [3] constructed a realistic dataset including CFA images and corresponding sRGB images. Each image is sub-sampled to reduce the influence of the noise.

2 Motivation

Although this dataset could be roughly used for realistic image denoising, it has many key problems:

Problem 1: The noise in sRGB space is no longer independent with the noise on other pixels, which only holds in the raw space. Thus, averaging the pixels on sRGB space is not reasonable.

Problem 2: Capturing 500 images take a long time, during which the scene should have no change in lighting conditions. So we cannot capture natural scenes or human which would be almost static in a short time while changing during a long time. For example, we need 500 seconds to take 500 images with a normal camera.

Problem 3: Take 500 images may generate shift on the spatial dimension.

Problem 4: The huge amount of heat generated by continuous taking photos will make the noise bigger

Problem 5: The physical property of each pixel has built-in attributes which make the noise in that pixel has a built-in tendency which cannot be averaged out by only using the 500 samples on that pixel.

Problem 6: Some special cameras, such as the camera in DJI UAVs, can only capture about 10 images each time and will automatically change the parameters in the next capturing. Hence, we cannot guarantee that the 500 images share the same camera settings.

Problem 7: The images in [3] are too small (of size 200×200) to contain any detailed information which can reflect the effectiveness of the demosaicking methods.

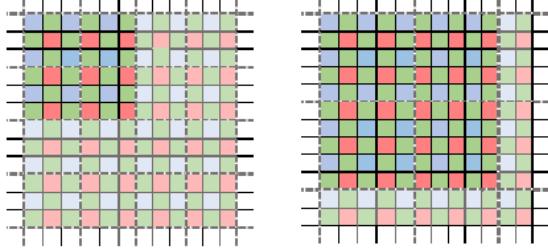


Figure 1: Averaging with odd windows, with $W = 1$ (left) and $W = 2$ (right), where the window size is $(2W + 1) \times (2W + 1)$.

3 Idea: Combining Spatial and Temporal Information

The above 6 problems are the major problems we need to solve if we want to take 500 pictures for each scene. The solution to **Problem 1** is very simple, we need to average the captured pictures in the raw space other the sRGB space. To solve the **Problems 2-7**, we can take few (e.g., 10) pictures instead of 500. To make up the decreasing of the temporal information, we can make use of the spatial information. That is, we can first average each CFA image via sub-sampling on the spatial direction and then average the 10 images along the temporal direction.

In Figure 1, we average the pixels of the $W \times W$ patch to form a red pixel, and average the pixels in the right $W \times W$ patch to form a green pixel, and average the pixels in the below $W \times W$ patch to form another green pixel, and average the pixels in the lower right $W \times W$ patch to form another blue pixel. The other patches are performed in a similar way.

4 Preliminary Results

We design a camera pipeline which includes (in sequential order)

- Black Light Subtraction;
- Spatial Sub-sampling;
- Temporal Averaging (10 images);

- White Balance;
- Demosaicking;
- Color Space Transfer;
- Gamma Correction;

The following example (Figure 2 and 3) is given by Hui Li. The Figure 2 is the sRGB image captured by the camera SONY A7 II when ISO=800. The Figure 3 is obtained by processing the original raw image without Color Space Transfer. The Demosaicking step is performed via the Deep Residual Network.

The following example (Figures 4 and 5) is given by Jun Xu. The Figure 4 is the sRGB image processed by the raw image captured by the camera SONY A7 II when ISO=12800. The Figure 5 is obtained by processing the original raw image with a simple camera pipeline.

One can see from the two examples that the noise in the raw images can be highly reduced. However, in some dark area, there still exists some color artifacts or noise.

5 Limitation of the Proposed Method

There are two major limitations:

- In some dark area, there still exists some color artifacts or noise. I think these problems can hardly be solved by adding more frames.
- The quality of the image generated by our pipeline is limited. This is because the pipeline is simple and rough.

6 Future Work

I think we can perform spatial sub-sampling and temporal averaging directly on sRGB images captured by cameras. In this way, we can construct “ground truth” images both for real image denoising and demosaicking problems.



Figure 2: The real noisy image processed by the raw image captured by SONY A7 II when ISO=800.



Figure 3: The averaged image obtained by spatial sub-sampling and temporal averaging.



Figure 4: The real noisy image processed by the raw image captured by SONY A7 II when ISO=12800.



Figure 5: The averaged image obtained by spatial sub-sampling and temporal averaging.

References

- [1] G. E Healey and R. Kondepudy. Radiometric CCD camera calibration and noise estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(3):267–276, 1994. [1](#)
- [2] S. Nam, Y. Hwang, Y. Matsushita, and S. J. Kim. A holistic approach to cross-channel image noise modeling and its application to image denoising. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1683–1691, 2016. [1](#)
- [3] Daniel Khashabi, Sebastian Nowozin, Jeremy Jancsary, and Andrew W Fitzgibbon. Joint demosaicing and denoising via learned nonparametric random fields. *IEEE Transactions on Image Processing*, 23(12):4968–4981, 2014. [1](#)