

生成噪声文章一：低照度下更精确的物理模型

来自北理工的硕士生，当年 CVPR 最佳论文展示奖。嗯，感觉确实还行，但也不至于最佳。不过作者确实很强，有扎实的数学基础。BIT 的本硕，之前去过剑桥和微软亚研院访问过，NB。

A Physics-based Noise Formation Model for Extreme Low-light Raw Denoising, CVPR 2020, KaiXuan Wei, BIT

Introduction & Related Works

低照度的摄像不可避免有问题：曝光短则噪声多、长则模糊和拖影发生概率高；其中一个解决方法 flash 也会导致颜色失准并且使用场景严格。

Large aperture incurs small depth of field, and is unavailable in smartphone cameras; long exposure can induce blur due to scene variations or camera motions; flash can cause color aberrations and is useful only for nearby objects.

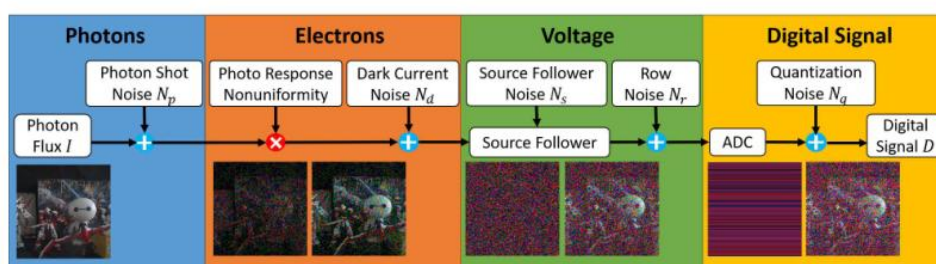
有两种方法：神经网络方法，用神经网络模拟噪声，但是需要训练需要干净的照片和有噪声的照片这个 pair，所以非常耗力耗时。而物理方法则取决于物理模型准不准确。

Such a deep learning approach generally requires a large amount of labelled training data that resembles low-light photographs in the real world

Notwithstanding the promising results, collecting sufficient real data with ground-truth labels to prevent overfitting is exceedingly expensive and time-consuming

模型建立

很常规，噪声有 shot noise 和 read noise，不过下面这张图挺不错的：



1. From Photon to Electrons

这个流程大部分是 shot noise，但是作者也做了说明，这部分还包括了：photo response nonuniformity 和 dark current noise。随着工艺的发展，这两个噪声降低了很多，并且将暗电流的波动放入 read noise 中。如果只考虑 shot noise，结果如下，和以往没有区别：

$$(I + N_p) \sim \mathcal{P}(I)$$

2. From Electrons to Voltage

这一步的噪声取决于：the circuit design and processing technology used。包括了：thermal noise, reset noise, source follower noise, banding pattern noise。统一用读出噪声表示，其中条带噪声先不管：

$$N_{read} = N_d + N_t + N_s$$

这里就是文章的关键了：以往的方法都是认为是高斯分布，但是统计发现有 long-tailed 现象，作者做了如下解释：

This can be attributed by the flicker and random telegraph signal components of source follower noise [25], or the dark spikes raised by dark current

因此作者没有用高斯分布，而是使用 TL 分布 (Tukey lambda)。而忽略掉的条带噪声，作者则是使用了高斯分布进行模拟。其中条带噪声分为 row 和 vertical，但是作者在实验中发现：vertical 的噪声基本可以没有（就是频率谱中没有横线）。

3. From Voltage to Digital Numbers

这一步就是量化噪声，和以前文献认知相同。

模型标定

需要两种类型的帧：Flat-filed frames 和 Bias frames，其实就是分别的亮暗帧。

1. 标定 shot noise

在 shot noise 部分，因为有增益，所以最后 $L = K * (I + N_p)$ 。首先需要量化增益 K 应该是多少。作者通过参考文献[32]的方法量化出来，方法是很经典的方法：

不同亮度（曝光时间）下的亮场图，对于每一张图，计算出均值、方差，有均值-方差这个 pairs，最后最小二乘法模拟出一条直线即可。

得出 K 之后，如何生成 shot noise？那就很简单了，直接根据泊松噪声分布生成就可以了。比如干净图中的像素值是 I，那么就生成分布是 P(I) 的值，最后乘以一个 K 就好了。

2. 标定非 shot noise

首先对于 Bias 而言，其频率谱如图。没有横线说明没有竖线条带，竖线则说明有 row noise：

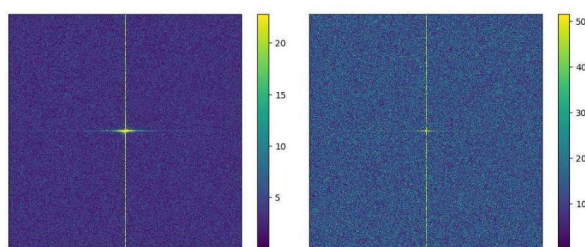


Figure 3: Centralized Fourier spectrum of bias frames captured by SonyA7S2 (*left*) and (*right*) NikonD850 cameras

标定 row noise 很朴素，直接对每一行求平均。作者使用 Shapiro-Wilk 检验，去查看各个 row noise 是否正态分布，结果是需要考虑的。所以 row noise 还有个标准差，记作 σ_r 。这里的标准针对所有的数据，也就是如果模拟噪声时，每一行的像素各自加上自己对应的 row noise，再加上该标准差的高斯噪声。

此时，剩余的噪声期望是 0，作者比较高斯分布、之前说的 TL 分布。评价指标是 R 方 (PPC C 图)，结果显示 TL 更好一点，其中分布的方差记作 σ_{TL} 。结果如下图所示：

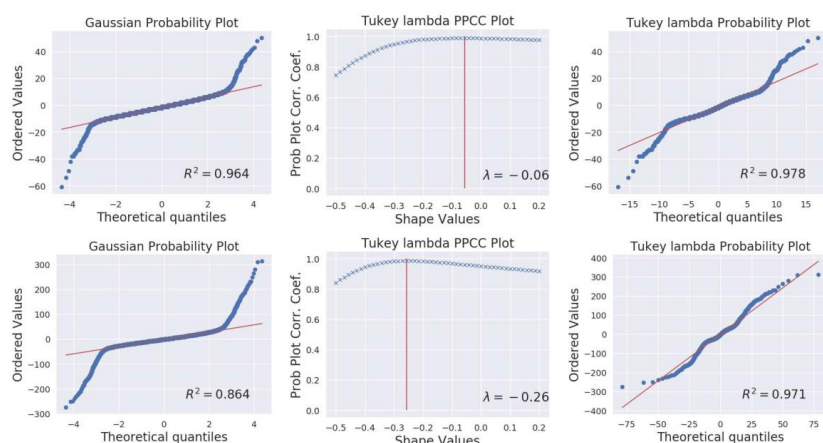


Figure 4: Distribution fitting of read noise for SonyA7S2 (*top*) and NikonD850 (*bottom*) cameras. *Left*: probability plot against the Gaussian distribution; *Middle*: Tukey lambda PPCC plot that determines the optimal λ (shown in red line); *Right*: probability plot against the Tukey Lambda distribution. A higher R^2 indicates a better fit. (Best viewed with zoom)

3. 自动化标定

我们得到了一组参数： K, σ_r, σ_{TL} ，这个对于一个 ISO 而言的，也就是不同的 ISO 这一组参数不相同。那干脆每个 ISO 都标定？太麻烦了！

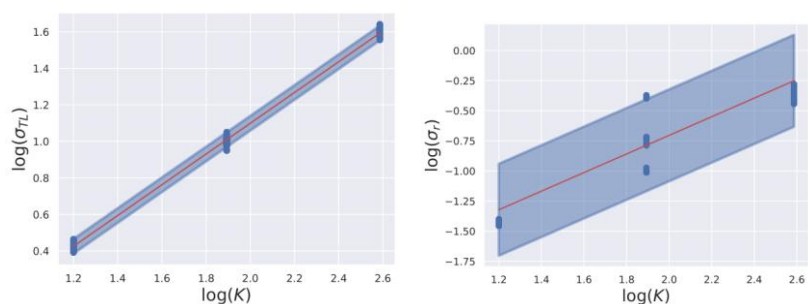
所以使用采样，只对几个 ISO 进行标定，然后建立参数之间的联合分布，然后通过最小二乘法

计算得出，假如不是提前设好的 ISO 时，应该是什么参数。

$$\begin{aligned}\log(K) &\sim U\left(\log(\hat{K}_{min}), \log(\hat{K}_{max})\right), \\ \log(\sigma_{TL}) | \log(K) &\sim \mathcal{N}(a_{TL} \log(K) + b_{TL}, \hat{\sigma}_{TL}), \quad (7) \\ \log(\sigma_r) | \log(K) &\sim \mathcal{N}(a_r \log(K) + b_r, \hat{\sigma}_r),\end{aligned}$$

其中 Kmin 和 Kmax 分别是在最小 ISO 和最大 ISO 得到的值。从第一个式子看出，认为 $\log(K)$ 和 ISO 成正比，此时有了一个新的 ISO 时，我们能直接得到 K。上述公式要训练的值是 a、b 和两个尖括号 sigma，这样有 K 之后，就能够得到想要的两个 sigma 值。

如下图所示，模拟出的是红线。而上下方的蓝线就是因为高斯分布（两个尖括号 sigma）有一定的误差。



参考文献

首先是论文很可，知乎上有一篇讲解很好（作者 BIT 的学弟），最后是 github 上有代码。

<https://www.zhuanlan.zhihu.com/p/356933763>