RENOIR - A Dataset for Real Low-Light Noise Image Reduction

Josue Anaya, Adrian Barbu

Department of Statistics, Florida State University, USA

Abstract

The application of noise reduction or image denoising is a very important topic in the field of computer vision and image processing. Many modern and popular state of the art image denoising algorithms are trained and evaluated using images with added artificial noise. These trained algorithms and their evaluations on synthetic data may lead to incorrect conclusions about their performances on real noise. In this paper we introduce a benchmark dataset of uncompressed color images corrupted by natural noise due to low-light conditions, together with spatially and intensity-aligned low noise images of the same scenes. The dataset contains over 100 scenes and more than 400 images, including both 16-bit RAW formatted images and 8-bit BMP pixel and intensity-aligned images from 2 digital cameras (Canon S90 and Canon T3i) and a mobile phone (Xiaomi Mi3). We also introduce a method for estimating the true noise level in each of our images, since even the low noise images contain a small amount of noise. Finally, we exemplify the use of our dataset by evaluating four denoising algorithms: Active Random Field, BM3D, Bilevel MRF optimization, and Multi-Layer Perceptron. We show that while the Multi-Layer Perceptron algorithm works as well as or even better than BM3D on synthetic noise, it does not do the same on our dataset.

Keywords: image denoising, denoising dataset, low light noise

*Corresponding author.
Email address: abarbu@stat.fsu.edu (Adrian Barbu)
URL: http://stat.fsu.edu/~abarbu/ (Adrian Barbu)
1. Introduction and Motivation

In the field of computer vision and computational photography, noise reduction is the application in which granular discrepancies found in images are removed. The task of performing noise reduction is synonymous with improvement in image quality. Many consumer cameras and mobile phones deal with the issues of low-light noise due to small sensor size and insufficient exposure. The issue of noise for a particular digital camera is so important that it is used as a valuable metric of the camera sensor and for determining how well the camera performs [Labs (2009)]. Another example of images that deal with noise due to limited acquisition time are Magnetic Resonance Images (MRI). Other important types of image modalities such as X-ray and CT (Computed Tomography) also suffer from noise artifacts due to insufficient exposure because of low radiation dose limits. While the image acquisition process is different in all of these examples, the reason for the noise is in most part the same. This is why the problem of low-light image noise reduction is studied and has led to a variety of different noise reduction methods.

In general most of the performance evaluations for these various noise reduction methods are done on small images contaminated with artificial noise (Gaussian, Poisson, salt and pepper, etc), which is artificially added to a clean image to obtain a noisy version. Measuring the performance of a noise reduction algorithm on small images corrupted by artificial noise might not give an accurate enough picture of the denoising performance of the algorithm on real digital camera or mobile images in low-light conditions. The nature of the noise in low-light camera images is more complex than just i.i.d., for example its variance depends on the image intensity [Foi et al. (2008)], so it would be desirable to obtain images naturally corrupted by low-light noise and their noise-free counterparts. For this purpose, we bring the following contributions:

- A dataset of color images naturally corrupted by low-light noise, taken with two digital cameras (Canon PowerShot S90, Canon EOS Rebel T3i) and a mobile phone camera (Xioami Mi3).
- A process for the collection of noisy and low-noise pixel-aligned images of the same scene.
- A method for aligning the intensity values of all the images of the same scene.
A technique for computing the noise level and Peak Signal-to-Noise Ratio (PSNR) \cite{teo1994} of the images in our dataset.

Different cameras produce different kinds of noise due to their sensor size, sensor type, and other factors on the imaging pipeline. A learning-based denoising method (e.g. \cite{barbu2009} or \cite{burger2012}) could be trained for a specific type of camera on noisy-clean image pairs for that specific camera, but it is not clear how well it would work on images from another camera. Trying to construct a dataset of various image pairs from different cameras may help determine which denoising method generalizes well over many different cameras, however it does not evaluate the full potential of a method on any one specific camera at various noise levels. We therefore selected to obtain an equal number of images from three cameras with different sensor sizes: one with a small sensor (Xiaomi Mi3), one with a slightly larger sensor (Canon S90) and one with a mid-size sensor (Canon T3i).

### 1.1. Related Works

Most digital cameras provide color images and it has been shown that the distribution of low-light camera noise is not Gaussian, but follows a more complex distribution with intensity-dependent variance \cite{foi2008, Luisier2011}. Although a variety of color image databases exist, such as \cite{ponomarnko2008}, the only image denoising database for benchmarking is \cite{estrada2010} (discussed in \cite{estrada2009}), which evaluates various denoising methods on color images corrupted by artificial Gaussian noise. The problem with artificial Gaussian noise is that a denoising method could be parameterized in such a way to give the best possible performance on such a simple noise model and still under-perform on images with real and more complex noise structure.

Work by both \cite{foi2008} and \cite{luisier2008} use a few real low-light noise and clean image pairs to describe the noise and intensity relationship. However, we are not aware of any public database or collection of images that have been corrupted by real low-light noise like the ones presented in our paper.

### 2. Acquisition of Natural Image Pairs

The dataset acquired in this paper consists low-light uncompressed natural images of multiple scenes. About four images per scene were acquired, where two images contain noise and the other two images contain very little noise. The presence of noise in the images is mainly due to the number of photons that are
received by the camera’s sensor and the amplification process, as shown in [Ishii et al. (2007)].

All the images in our dataset are of static scenes and are acquired under low-light conditions using the following "sandwich" procedure:

- A low-noise image is obtained with low light sensitivity (ISO 100) and long exposure time. This will be the reference image.
- One or two noisy images are then obtained with increased light sensitivity and reduced exposure time.
- Finally, another low noise image is taken with the same parameters as the reference image. This will be the clean image.

The two low-noise (clean) images are acquired at the beginning and at the end of the sequence, while the one or two noisy images are shot in between. This is done to evaluate the quality of the whole acquisition process for that particular scene. This process is very similar to the process discussed in [Lui et al. (2008)] which used pair images that were taken with flash. The problem with taking the images with flash is that the illumination can change the scene. Moreover, they do not perform any brightness and contrast alignment on their images. For our dataset if there was some motion or lighting change during acquisition, the two clean images will be sufficiently different, as measured by the PSNR, obtained as described in section 2.3. In fact, we discarded the scenes with PSNR of the clean images less than 34.

The actual acquisition parameters for each camera are presented in Table 1.

| Camera | ISO  | Time(s) | ISO  | Time(s) |
|--------|------|---------|------|---------|
| Mi3    | 100  | auto    | 1600 or 3200 | auto    |
| S90    | 100  | 3.2     | 640 or 1000  | auto    |
| T3i    | 100  | auto    | 3200 or 6400 | auto    |

Using the Canon Developer Tool Kit for the Canon S90 and the EOS Utility for the Canon Rebel T3i we were able to program the automatic collection of the four images while trying to preserve the static scene in the images by not moving or refocusing the camera. The sandwich approach that we used to obtain our images also helped insure that the only visual difference between the images of a scene was simply due to noise. All of the images were collected and saved in RAW format (CR2 for Canon). The Mi2Raw Camera app was used to capture the RAW images for the Xiaomi Mi3 (in DNG format).
Figure 1: An example of a clean and noisy image pair as well as their corresponding blue channel. The noise present is a result of the low-light environment. The images were taken using a Canon PowerShot S90.

An example of one of the images in the dataset can be seen in Figure 1. In the end we collected 51 scenes for the S90, 40 for the T3i, and another 40 for the Mi3.

The image denoising database in Estrada et al. (2010) contains 300 noisy images at 5 noise levels ($\sigma = 5, 10, 15, 25, 35$) for a total of 1500 images. The dimensionality of these images is 481 by 321. The dimensionality of the images for just the S90 images is 3684 by 2760 while the images from the other cameras are even larger, as shown in Table 2. Although our image database contains far fewer noisy images, our images contain about 60 times more pixels and therefore more patch variability for studying noise models from just one of the three cameras.

Many various denoising methods Burger et al. (2012); Barbu (2009) train models from the noisy-clean image pairs that are supposed to generalize well to future noisy images. For this reason and for evaluation in general it is very important to maintain a careful construction of these noisy-clean image pairs and to
Table 2: Description of the dataset and size

| Camera | Image Size | Sensor Size (mm) | Scenes | Noisy Images | Clean Images |
|--------|------------|------------------|--------|--------------|--------------|
|        | RAW        |                  |        | σ PSNR       | σ PSNR       |
|        | Sensor     |                  |        | Min. Avg. Max.| Min. Avg. Max.|
| S90    | 3684×2760  | 7.4×5.6          | 51     | 18.249 17.429 26.187 33.388 | 3.067 35.032 38.698 43.425 |
| T3i    | 5202×3465  | 22.3×14.9        | 40     | 11.714 18.938 27.442 35.256 | 2.568 34.983 40.426 48.128 |
| Mi3    | 4208×3120  | 4.69×3.52        | 40     | 19.227 12.751 23.492 36.678 | 3.712 33.495 37.093 45.252 |

have many examples for a representative dataset. The difficulty in constructing such pairs is why artificial noise is used in practice.

2.1. Mobile camera difficulties

In trying to collect images for our dataset we decided on also collecting mobile phone camera images. In doing so we ran into a variety of difficulties, some of which can be seen in Figure 2.

The first difficulty that arose was collecting the RAW images of these phone cameras. Only a few mobile phones collect true RAW images (data dump directly from the sensor). For example, the Iphone can have an application installed that will allow it to take RAW images, however these RAW images are not in fact truly RAW because the sensor data has gone through some unknown post processing. Therefore, only some of the most recent phones that have been allowed by the device manufacturer can truly collect RAW images.

The second difficulty that we found when trying to collect mobile phone images came in the control over certain image acquisition parameters such as the exposure time and ISO values. These mobile phone cameras already have a very small sensor, so when we tried to use an LG Google Nexus 5 with the FV-5 camera application to capture RAW images, the limits of control over settings like the exposure time and ISO, and its tiny sensor size led to many scenes failing our ‘sandwich’ procedure selection benchmark ( did not have sufficient amount of light needed for the PSNR to be around 35 for the reference and clean image.) We also noted for a phone like the LG Google Nexus 5 a non-linearity relationship issue in the brightness alignment procedure (this could also be due to an insufficient amount of light.)

The final difficulty we experienced came in the form of tools to help maintain a static scene. With the other cameras we used a tripod and were able to program the automatic acquisition of the scene. With the mobile phone camera we had to use a small phone tripod and a bluetooth mouse to preserve the static scene when taking the images manually.
Figure 2: Examples of alignment issues observed on scatter plots of the intensity difference between the aligned reference image and noisy image vs reference image intensity. The horizontal line in the plots are 95% noise bounds for images with at least PSNR = 35. Left: an example of movement during the ’sandwich’ procedure. Middle: an example of light saturation during the ’sandwich’ procedure. Right: an example of an image with too much noise and in need of some nonlinear transformation for proper brightness alignment.

Settling on the Xiaomi Mi3 phone we collected 6 images per scene. The first two images were both low-noise images and these images were averaged and set as the reference image in the alignment process. Similarly, the last two images were also both low-noise images and the last two images as well were averaged and used in the overall PSNR computation of the ’sandwich’ procedure. If any movement or saturation was detected the images were cropped appropriately post alignment. In the end many of the scenes for the Mi3 were cropped, but all are static with PSNR around 35 or more.

2.2. Main Assumptions and Notations

In this section we present the main assumptions that form the basis of the acquisition procedure, intensity alignment, and noise level estimation.

The following notations will be used in this paper:

- $R, I^r$ – the reference image
- $C, I^c$ – the clean image
- $N, I^n$ – the noisy image
- $X$ – one of the clean or noisy images
- $GT, I^{GT}$ – the unknown ground truth image
- $\epsilon, \epsilon_r, \epsilon_c$ – random variables for noise in the low-noise images
- $\epsilon_n$ – random variable for noise the noisy images
- $\sigma^2(X) = \text{var}(X)$ the variance of a random variable $X$

It is shown in [Lui et al., 2008] that the noise in the digital camera images has short range correlations. However, we will ignore these correlations and assume that the two low-noise images $I^r$ (reference) and $I^c$ (clean) as well as the noisy image(s) $I^n$ (acquired with the ”sandwich” procedure from Section 2) are all noisy versions of a common (unknown) ground truth image $I^{GT}$, corrupted by independent noise. We will see in experiments that our estimation method based on this assumption works very well in estimating the noise level in images. Note that the reference and clean images have low amounts of noise because many photons have been accumulated on the sensor during the long exposure time. We assume that the reference and clean images have the same noise distribution since the two images are of the same static scene with the same ISO level and exposure time. Thus,

$$I^n(x, y) = I^{GT}(x, y) + \epsilon_n(x, y)$$
$$I^r(x, y) = I^{GT}(x, y) + \epsilon_r(x, y)$$
$$I^c(x, y) = I^{GT}(x, y) + \epsilon_c(x, y)$$ (1)

2.3. Intensity Alignment

The dataset construction went beyond just the acquisition of the images. For the purposes of properly aligning the pixel intensities of the image pairs we developed a new form of brightness adjustment that mapped our RAW images to an 8-bit uncompressed format.

The reference image was first mapped from 16-bit to 8-bit as follows. We computed the cumulative distribution of the 16-bit pixel intensities of the RAW reference image and constructed a linear scaling of the RAW reference image that sets the 99th percentile value to the intensity value 230 in the 8-bit image. Thus 1% of the pixels are mapped to intensities above 230, and even fewer will be saturated to value 255. We chose the value 230 so that most of the noisy images will not have much saturation after alignment with the reference image.

Each of the other images of the same scene is at the same time reduced to 8-bit and aligned to the 8-bit reference image by finding a linear mapping specified by parameter $\alpha$ such that if $I$ is the 16-bit image, the 8-bit aligned image is obtained from $\alpha I$ after its values larger than 255 or less than 0 are truncated. For better accuracy, instead of working with the two images $I$ and $R$, we use blurred versions $\tilde{I}$ and $\tilde{R}$ obtained by convolution with a Gaussian kernel with $\sigma = 5$ to estimate
the intensity alignment parameter $\alpha$. This way the level of noise is reduced. To avoid biases obtained near intensity discontinuities, the alignment parameter is computed based on the low gradient pixels $M = \{i, |\nabla \tilde{R}(i)| < 1\}$.

The parameter $\alpha$ is found to minimize

$$E(\alpha) = \sum_{i \in M} (\tilde{R}(i) - \max[\min(\alpha \bar{I}(i), 255), 0])^2$$

This is done by coordinate optimization using the Golden section search in one dimension [Press (2007)] optimizing on $\alpha$ until convergence. The parameter $\alpha$ obtained for the mapping is robust to outliers. Figure 3 shows an example of the green channel for a particular image in the dataset before and after alignment is performed. Figure 4 shows an example of the mapping between the pixels $M$ of $\tilde{I}$ and $\tilde{R}$. We obtained plots like Figure 4 for all the images in the dataset as a way of diagnosing any misalignment or nonlinear correspondence between clean and noisy images. Note that in Figure 4 the middle and right plots show the pixel to pixel difference per intensity. The dark dashed horizontal line in the middle and right plots are 95% noise bounds for clean images with at least PSNR = 35.

2.4. Noise Estimation

As stated previously the amount of noise present in the dataset is due to the sensor and the amplification process. The fact that not all of the images were taken in the same environment under the same camera settings means that we have a wide variety of noise in our images. The fact that we are not dealing with artificial noise also means that we do not know beforehand what will be the noise variance $\sigma^2$. Thankfully our ”sandwich” procedure for image acquisition, as influenced by
Figure 4: Scatter plots of the pixel intensity correspondence between a reference image and its noisy counterpart. Left: the correspondence between the red channel of the 8-bit reference image and the red channel for the 16-bit noisy image. The line shows the estimated linear mapping to align the noisy image to the reference image. Middle: the difference between corresponding pixel intensities of the reference 8-bit and aligned noisy 8-bit image vs reference image intensities for all three color channels. Right: the difference between corresponding pixel intensities between the reference 8-bit and the aligned clean 8-bit image vs reference image intensities.

Healey and Kondepudy (1994); Lui et al. (2008), allows us to estimate the noise level for any one of our images.

We will use the fact that if two random variables $A, B$ are independent, then $\text{var}(A - B) = \text{var}(A) + \text{var}(B)$, or in other words $\sigma^2(A - B) = \sigma^2(A) + \sigma^2(B)$ where $\text{var}(A), \sigma(A)$ are the variance and standard deviation of $A$ respectively. Then from equation (1) we get

$$\sigma^2(I^r - I^c) = \text{var}(\epsilon_r - \epsilon_c) = \text{var}(\epsilon_r) + \text{var}(\epsilon_c) = 2\sigma^2(\epsilon)$$

from the independence of $\epsilon_r$ and $\epsilon_c$ and the fact that $\epsilon_r$ and $\epsilon_c$ are identically distributed (so we can represent them as $\epsilon$). We obtain the estimation of the noise level in the clean and reference images:

$$\sigma^2(I^r - I^{GT}) = \sigma^2(I^c - I^{GT}) = \sigma^2(\epsilon) = \frac{1}{2}\sigma^2(I^r - I^c) \quad (2)$$

For the noisy images we use

$$\sigma^2(I^n - I^r) = \text{var}(\epsilon_n - \epsilon_r) = \sigma^2(\epsilon_n) + \sigma^2(\epsilon_r)$$

to obtain the estimation of the noise level as

$$\sigma^2(\epsilon_n) = \sigma^2(I^n - I^{GT}) = \sigma^2(I^n - I^r) - \frac{1}{2}\sigma^2(I^r - I^c) \quad (3)$$
If we want to use the best estimate of the GT, which is \( I^a = \frac{I^r + I^c}{2} \), then we have an alternative formula for the noise level in the noisy images

\[
\sigma^2(\epsilon_n) = \text{var}(I^n - I^{GT}) = \text{var}(I^n - I^a) - \frac{1}{4} \text{var}(I^r - I^c)
\] (4)

We can use equations (2) and (3) to estimate the true noise level for any image in our dataset. These noise levels can be computed globally for the whole image or locally on a patch basis.

3. Dataset Information

Aside from estimating the noise level in every image, we also quantified the image fidelity across the image batches using various metrics such as PSNR [Teo and Heeger (1994)], SSIM [Wang et al. (2004)], and VSNR [Chandler and Hemami (2007)]. In particular we modified the PSNR measurement by incorporating our estimate of the noise from (3) as opposed to using the standard noise estimate from the difference image between a clean and noisy image pair. Although there exist specialized metrics for low-light conditions such as [Alter et al. (2006)] we decided to use measures that are the most prevalent and common in practice.

Table 2 lists some specific characteristics about the various cameras and their images in the dataset. Note that the \( \sigma \) in Table 2 comes from the estimates from equations (2) and (3). Figure 5 shows the distribution of noise levels for the noisy and clean images for each camera. Figure 6 shows box-plots of the variation in PSNR and noise levels for each camera.

Figure 5: Frequency distributions of various noise levels for the noisy and clean images obtained from the S90, T3i, and Mi3 cameras respectively.
4. Experiments

In this section we will be evaluating our noise estimation procedure and various denoising algorithms. With regards to the evaluation of noise estimation we will be comparing our noise estimation procedure to the standard noise estimate (standard deviation of the difference image) for both synthetic and real noise data. We will also be comparing our noise estimation framework to the Poisson-Gaussian noise model presented in [Foi et al. (2008)]. Afterwards we will examine the denoising performances of four algorithms on our dataset using three image fidelity metrics.

4.1. Evaluation of Noise Estimation Using Artificial Noise

To evaluate our noise estimation method we chose ten 16-bit RAW reference images from the three digital cameras and used them as ground truth images \( I^{GT} \) for constructing artificial sequences from them. We then used our alignment method as described in [2, 3] to construct an 8-bit version of \( I^{GT} \). We then generated \( I^r, I^n, \) and \( I^c \) by adding artificial Gaussian noise to the 16-bit \( I^{GT} \). For 16-bit \( I^r \) and \( I^c \) we added \( \sigma = \frac{3}{\gamma} \) amount of noise where \( \gamma \) is the multiplications factor to map the 16-bit \( I^{GT} \) to an 8-bit \( I^{GT} \). A 16-bit \( I^n \) was generated using \( \sigma = \frac{10}{7} \). This way the standard deviation of the difference from the 8-bit \( I^{GT} \) to the 8-bit \( I^r \) (or \( I^c \)) will be 3 and to the 8-bit \( I^n \) it will be 10. We then performed our standard alignment on \( I^r, I^n, \) and \( I^c \) to map them over to 8-bit images obtaining parameters \( \gamma', \alpha_1, \alpha_2 \) as illustrated in Figure 7.

To estimate the true value of noise for \( I^r, I^n, \) and \( I^c \) we computed the standard deviation of the difference image between each of them and \( I^{GT} \) (all in 8-bit

![Figure 6: Variation of PSNR and \( \sigma \) values for noisy and clean images for each camera.](image)

12
Figure 7: The process for constructing the proper reference, clean, noisy, and ground truth images necessary for the noise estimation evaluation. The values of $\gamma$, $\alpha_1$, and $\alpha_2$ represent the usual alignment of those respective images from 16-bit to 8-bit as defined in section 2.3.

versions). We then compared the standard estimate of noise as $\sigma(I^n - I^c)$ and $\sigma(I^c - I^r)$ with the noise estimation of $I^c$ and $I^n$ using our method explained in Section 2.4.

Figure 8: The relative error in estimating noisy and clean images.

Figure 8 shows the relative error (defined as error divided by noise level) of estimating the noise level for both $I^n$ and $I^c$. When it came to estimating the noise level $I^n$ our method of estimation kept the relative error to below 0.5%, while the standard method of estimating the noise level had a relative error around 5%.

This experiment shows that the alignment method together with the noise level estimation method work well together and obtain a very small error in estimating the noise level, at least on data with artificial noise.
4.2. Evaluation of Noise Estimation Using Real Noise

To further investigate how well the assumptions we made in Section 2.2 about the noise hold, we acquired a special scene with the S90 camera. The scene was of a constant intensity surface in low-light settings. Using our intensity alignment methodology, instead of mapping our clean image from the 99th quantile to intensity 230; we mapped the median to intensity 128. Using this mapping we then aligned the other two noisy and the clean image using the Golden section method, as described in Section 2.3. Figure 9 shows the alignments of the calibration dataset as well as a histogram of pixel difference between the reference image and the other images in the calibration dataset.

![Figure 9: Analysis of the calibration images. Left: the intensity histograms of the green channels of the calibration images. Right: the distribution of the intensity difference between the reference image and the various other images in the calibration dataset.](image)

Since we know that the $I^{GT}$ was constant since the scene contained a constant intensity surface, we can immediately obtain a true value for $\sigma^2$ for each image by directly computing the intensity variance in the image. However, to account for smoothly changing illumination, we constructed a GT version for each image by Gaussian blurring it with a large spatial kernel ($\sigma = 20$) and then calculated the noise level as the variance of the difference image between the original image and its smoothed version. We then looked to see if the standard estimate of using the difference image between the reference image and the other calibration images provided similar results to those we obtained using our methodology from equations (2) and (3). Analysis of the estimated noise levels for the three image channels and the overall estimate are summarized with boxplots in Figure 10.

As Figure 10 shows our estimated $\sigma$ values are less biased and have smaller variance than the standard estimation of $\sigma$ from the difference images. The average relative error for our method of estimation is 1.58% and for the standard method of estimation is 36.22%. The results that we obtained for this evaluation...
are in line with the results we obtained for noise estimation for images with artificial noise. Thus our investigation gives us enough confidence in our estimation going forward. Consequently, the noise estimation described in Section 2.4 will be used as our noise estimation method for all of the images in our dataset and for estimating the PSNR of the denoised images.

### 4.3. Evaluation of the Poisson-Gaussian Noise model

As stated previously the noise depends on the image intensity. The Poisson-Gaussian noise model presented in Foi et al. (2008) provides for a maximum likelihood estimate of the noise curve. Equations (2), (3), and (4) return a single value estimate for the noise if we allow \( I^r \), \( I^c \), and \( I^n \) to be the usual entire image. Our noise estimation framework can also allow for corresponding subsets of \( I^r \), \( I^c \), and \( I^n \) to be used in local neighborhood noise estimation. It is through this result that we can calculate the noise variance in an image for each intensity level and compare it with the maximum likelihood estimate from Foi et al. (2008).

A special scene was acquired using a uniform background with a smoothly changing intensity and our "sandwich" procedure. We converted the images to gray-scale to be able to compute the model from Foi et al. (2008) and computed many noise curves:

- The image pair model is obtained using equations (2) and (4).
- The Poisson mixture model is obtained from the noisy image as the maximum likelihood estimate of the Poisson-Gaussian mixture model using the methods presented in Foi et al. (2008).
- The standard image pair model was the noise curve obtained using the standard noise estimation from the difference between the reference and the
noisy image.

- The blurred reference model was obtained by adding a lot of blur to the reference image to better approximate the smooth ground truth image and computed the standard noise curve from the difference between this blurred reference image and the noisy image.

![Figure 11: Left: The noise curve of our pair image model and the Foi Poisson-Gaussian Mixture model. Right: The noise curves for various noise estimation models using image pairs and the Foi Poisson-Gaussian Mixture model.](image)

The various curves are shown in Figure 11. In Figure 11 left are also shown the data points from which the image pair model and the Poisson mixture model were obtained. In Figure 11 right are also shown the data points from which the blurred reference model (blue) and the standard image pair model (red) were obtained.

It can be immediately noted that the Poisson-Gaussian noise model is consistently below the other three curves, which are quite close to each other. At high intensities (around 0.9 on the normalized scale), Foi’s estimate using the Poisson-Gaussian noise model is underestimating the noise pair curve estimate by about 40%. However, looking closer at the estimated noise curve and the true noise curve one can see that they share relatively the same structure and shape.

Foi’s Poisson-Gaussian noise model is using just the noisy image to infer the ground truth image. With our image pair model we have both clean and noisy image pairs and are able to infer more accurately the ground truth image. This is why in Figure 11 we can see all the pair image models are grouped closer together than the Poisson-Gaussian mixture model. Also, note that this evaluation was
on a special scene of a uniform background of continuously changing intensity and no edges. Foi’s Poisson-Gaussian noise model would have more difficulty in estimating the noise curve in images with a lot of edges or a lot of textures.

4.4. Evaluation of Denoising Algorithms

In this section we use our dataset to evaluate four popular image denoising algorithms: the Active Random Field (ARF) [Barbu (2009)], BM3D [Dabov et al. (2007)], Bilevel optimization (opt-MRF) [Chen et al. (2013)] and Multi Layer Perceptron (MLP) [Burger et al. (2012)]. These algorithms were selected because they are reasonably fast to work with our large images and have code available online. Each of these methods depends on a noise level parameter $\sigma$. The methods were evaluated for a number of values of the noise level $\sigma$ and the best results were reported for each method. We tested the ARF filters (which can be found at [http://www.stat.fsu.edu/~abarbu/ARF/demo.zip](http://www.stat.fsu.edu/~abarbu/ARF/demo.zip)) that were trained using Gaussian noise (in particular the trained filters for $\sigma = 10, 15, 20, 25$ and 50) and using four iterations. A special version of the BM3D algorithm meant for color image denoising (which can be found at [http://www.cs.tut.fi/~foi/GCF-BM3D/BM3D.zip](http://www.cs.tut.fi/~foi/GCF-BM3D/BM3D.zip)) was used on the noisy images. For BM3D we evaluated the algorithm’s performance at $\sigma = 5, 10, 15, 20, 25, \text{ and } 50$. For opt-MRF we used the Gaussian trained filters ($\sigma = 15$ and 25) and a maximum limit of 30 iterations for the optimization (the filters and code can be found at [http://pan.baidu.com/share/link?shareid=1707746373&uk=2974149445](http://pan.baidu.com/share/link?shareid=1707746373&uk=2974149445)). Finally, we used MLPs trained on Gaussian filters (which can be found at [http://people.tuebingen.mpg.de/burger/neural_denoising/](http://people.tuebingen.mpg.de/burger/neural_denoising/)) to denoise our images. In particular we used filters for $\sigma = 10, 25, 35, 50, \text{ and } 75$. For the ARF, opt-MRF, and MLP algorithms the images were denoised in the YUV color space for better performance.

In Table 3 are shown the denoising results of the various methods on the three cameras. We computed the PSNR, SSIM, and VSNR values between the denoised and the best GT estimate which is the average of the two clean images. Note the high values prior to denoising given by the SSIM. We believe the high measurement values are due to our alignment procedure preserving a lot of the image structure while still being noisy.

The best results obtained for the ARF, opt-MRF, and MLP methods occurred with a $\sigma = 25$ filter while the BM3D provided its best results with a $\sigma = 50$ filter. The results for Table 3 show that the ARF, opt-MRF, and MLP methods performed about the same, while BM3D performed the best on all three cameras. In particular when examining the performance between MLP and BM3D for real
Table 3: Denoising results

| Camera | Before Denoising | ARF | BM3D | opt-MRF | MLP |
|--------|------------------|-----|------|---------|-----|
|        |                  |     |      |         |     |
| PSNR   |                  |     |      |         |     |
| Mi3    | 23.492           | 30.918 | 32.347 | 31.641 | 31.230 |
| S90    | 26.187           | 33.797 | 36.752 | 34.983 | 34.073 |
| T3i    | 27.442           | 36.550 | 39.966 | 38.646 | 37.584 |
| Average| 25.707           | 33.755 | 36.355 | 35.090 | 34.296 |
| SSIM   |                  |     |      |         |     |
| Mi3    | 0.989            | 0.972 | 0.982 | 0.964   | 0.929 |
| S90    | 0.988            | 0.959 | 0.979 | 0.958   | 0.920 |
| T3i    | 0.991            | 0.993 | 0.994 | 0.993   | 0.933 |
| Average| 0.989            | 0.981 | 0.985 | 0.972   | 0.927 |
| VSNR   |                  |     |      |         |     |
| Mi3    | 17.746           | 22.387 | 24.820 | 22.521 | 24.132 |
| S90    | 23.789           | 26.769 | 28.635 | 27.357 | 27.255 |
| T3i    | 22.318           | 28.567 | 30.481 | 29.803 | 29.429 |
| Average| 21.284           | 25.908 | 27.979 | 26.560 | 26.605 |

noisy images these results do not lead to the same conclusions as in [Burger et al. (2012)] where MLP was comparable to or even slightly outperformed BM3D. In the case of the SSIM measure all four denoising methods struggled with the Mi3 and S90 camera, while the MLP underperformed on all three cameras.

5. Conclusions

In this paper we introduced a dataset of images containing real noise due to low-light settings and acquired from two digital cameras and a mobile phone. Additionally, we developed a method for obtaining pixel-aligned RAW images of low and high noise, and intensity-aligned BMP images so that proper studying of the images and its noise need not be only done in RAW format. We also presented a technique to calculate the PSNR of an image without a ground truth. Finally, we conducted and performed extensive evaluations of our noise estimation and our alignment procedure to make sure that the difference between the noisy and clean images is just noise.

We tested our dataset on four denoising algorithms ARF, BM3D, opt-MRF and MLP. We were then able to calculate and measure the noise levels in the denoised images using a variety of different methods such as PSNR, VSNR, and SSIM. Note that these methods were trained or tuned on images corrupted by artificial Gaussian noise. Some of these methodologies (ARF opt-MRF, and MLP) and many other recent state-of-the-art denoising methods such as: CSF [Schmidt and...
Roth (2014), LSSC Mairal et al. (2009), and RTF Jancsary et al. (2012) learn the noise structure from the noisy-clean image pairs. These methods could in fact perform even better for denoising low light images if trained on our dataset. With so many different denoising methods having been developed or currently in development, our dataset allows for proper analysis of these tools, and for the quantitative evaluation of noise models for digital and mobile phone cameras.

Acknowledgements

This work was supported in part by DARPA MSEE grant FA 8650-11-1-7149.

References

Alter, F., Matsushita, Y., Tang, X., 2006. An intensity similarity measure in low-light conditions. In: ECCV. pp. 267–80.

Barbu, A., 2009. Training an active random field for real-time image denoising. Image Processing, IEEE Transactions on 18 (11), 2451–2462.

Burger, H. C., Schuler, C. J., Hamerling, S., 2012. Image denoising: Can plain neural networks compete with bm3d? In: CVPR. pp. 2392 – 2399.

Chandler, D. M., Hemami, S. S., 2007. Vsnr: A wavelet-based visual signal-to-noise ratio for natural images 16, 2284 –2298.

Chen, Y., Pock, T., Ranftl, R., Bischof, H., 2013. Revisiting loss-specific training of filter-based mrfs for image restoration. In: German Conference Pattern Recognition. pp. 271–281.

Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K., 2007. Image denoising by sparse 3-d transform-domain collaborative filtering. Image Processing, IEEE Transactions on 16 (8), 2080–2095.

Estrada, F., Fleet, D., Jepson, A., 2009. Stochastic image denoising.

Estrada, F., Fleet, D., Jepson, A., 2010. Image denoising benchmark. http://www.cs.utoronto.ca/~strider/Denoise/Benchmark/ [Online; accessed 15-April-2014].
Foi, A., Trimeche, M., Katkovnik, V., Egiazarian, K., 2008. Practical poissonian-gaussian noise modeling and fitting for single-image raw-data. Image Processing, IEEE Transactions on 17 (10), 1737–1754.

Healey, G. E., Kondepudy, R., 1994. Radiometric ccd camera calibration and noise estimation 16, 267–276.

Ishii, Y., Saito, T., Komatsu, T., 2007. Denoising via nonlinear image decomposition for a digital color camera. In: ICIP. Vol. 1. IEEE, pp. I–309.

Jancsary, J., Nowozin, S., Rother, C., 2012. Loss-specific training of nonparametric image restoration models: A new state of the art. In: European Conf. Computer Vision. pp. 112–125.

Labs, D., 2009. Dxomark sensor scores. [Online; accessed 4-April-2014].

Lui, C., Szeliski, R., Kang, S. B., Zitnick, C. L., Freeman, W. T., 2008. Automatic estimation and removal of noise from a single image 30, 299–314.

Luisier, F., Thierry, B., Unser, M., 2011. Image denoising in mixed poisson-gaussian noise. Image Processing, IEEE Transactions on 20, 696–708.

Mairal, J., Bach, F., Ponce, J., Sapiro, G., Zisserman, A., 2009. Non-local sparse models for image restoration. In: IEEE Int’l Conf. Computer Vision. pp. 2272–2279.

Ponomarnko, N., Lukin, V. L., Egiazarian, K. O., Astola, J., Carli, M., Battisti, F., 2008. Color image database for evaluation of image quality metrics. In: International Workshop on Multimedia Signal Processing. IEEE.

Press, W. H., 2007. Numerical recipes 3rd edition: The art of scientific computing. Cambridge University Press.

Schmidt, U., Roth, S., 2014. Shrinkage fields for effective image restoration. In: IEEE Int’l Conf. Computer Vision and Pattern Recognition.

Teo, P. C., Heeger, D. J., 1994. Perceptual image distortion. In: International Symposium on Electronic Imaging: Science and Technology. pp. 127–141.
Wang, Z., Bovik, A. C., Sheikh, H. R., Simoncelli, E. P., 2004. Image quality assessment: from error visibility to structural similarity 13, 600–612.

Josue Anaya received the B.Sc. and M.Sc. degrees in statistics from Florida State University, Tallahassee, in 2011 and 2013 respectively. His research interests include statistical modeling, computational photography, computational imaging, large data analytics, and neural networks. He is a current Ph.D candidate expecting to obtain his Ph.D from Florida State University in 2016.

Adrian Barbu received his B.Sc. degree from University of Bucharest, Romania, in 1995, a Ph.D in Mathematics from Ohio State University in 2000 and a Ph.D in Computer Science from UCLA in 2005. From 2005 to 2007 he was a research scientist and later project manager in Siemens Corporate Research, working in medical imaging. He received the 2011 Thomas A. Edison Patent Award with his co-authors for their work on Marginal Space Learning. From 2007 he joined the Statistics department at Florida State University, first as assistant professor, and since 2013 as associate professor. His research interests are in computer vision, machine learning and medical imaging.