

A Novel Image Noise Level Function Estimation Approach using Camera Response Function Constraint

Heng Yao

School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, 200093 Shanghai, China

Abstract. In this paper, we propose a noise level evaluation method for real captured photos. Different from conventional noise removal methods that assume noise follow a simply additive Gaussian distribution, noise distribution in our method is supposed to be a more sophisticated intensity-dependent distribution, which has a better fit with actual noise model. Follow the definition of noise level function (NLF) which represents the variation of the standard deviation of noise with respect to image intensity. After exposing the close relationship between NLF and camera response function (CRF), we fit the curve of NLF with the constraints imposed by the shape of CRF.

Index Terms— image noise estimation, noise level function, camera response function

1 Introduction

For noise contaminated photos, random noise is mainly caused by quantum effects, thermal fluctuations and dark current leakage. Although hardware manufacturers have been trying to suppress random noise, the problem is still inevitable. Estimating the noise level from a single image is a significant task in many computer vision algorithms, especially for de-noising and forensic applications. Most algorithms [1-2] use the additive white Gaussian noise (AWGN) model, and generate noisy images by adding white noise manually into noise-free images in the experiments. However, the AWGN conjecture may not hold for real-life digital photographs because the actual CMOS/CCD sensor noise is strongly dependent on the light intensity, and the forgery makers are unlikely to deliberately add noise to lower the visual quality of the fake images. Based on this consideration, Liu et al. [3] define a noise level function (NLF) with respect to image intensity. They collect a sample set representing spatial average and variation, and find the lower envelope of the samples. A Bayesian framework is used to optimize the fitting. The method is later extended to an application to noise removal [4]. In this paper, we address the issue of accurate evaluation of the noise level from an image with moderate noise level using an incomplete sample set.

2 Model of Noise Level Functions

In this section, we explore the correlation between the camera response function (CRF) and the corresponding noise level function (NLF). An NLF model is then

established, which will be used to estimate NLF in Section 3.

2.1 Camera response function

Many image processing algorithms assume that the observed image intensity is linearly proportional to the scene radiance recorded by the camera sensor. However it is generally not the case. The camera response function (also termed the radiometric response function) $f(\cdot)$ is defined to describe the non-linear mapping between the scene radiance R and the measured intensity I in an image:

$$I = f(R) \quad (1)$$

Assuming that f is continuous and monotonic, it can be inverted to obtain the inverse response function, denoted $g(\cdot)$. Neglecting the errors due to discretization, R can be obtained by $R = g(I)$. Since only the observed output intensity I is available, most CRF estimation approaches attempt to find the inverse CRF g instead of f .

2.2 Noise level function

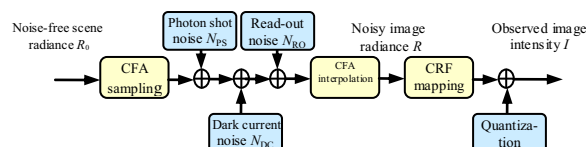


Figure.1. Imaging process of transferring radiance photons into intensity bits.

As pointed out in [5], noise produced in a digital camera is not simply additive, but strongly dependent on the

image intensity. To describe the relationship between noise and image intensity, the noise level function (NLF) is defined as the variation of the standard deviation of noise with respect to image intensity [3]. Fig. 1 illustrates the imaging process of transferring radiance photons into intensity bits. Scene radiance passes through the lens and is contaminated by several kinds of noise prior to the transformation into intensity bits. Let R_0 , R and I denote noise-free scene radiance, noisy radiance and observed intensity respectively. Intensity of the noise-free image is denoted as I_0 , which cannot be obtained from a single image. There are four main types of noise [6]: photon shot noise, dark current noise, read-out noise and quantization noise, denoted N_{PS} , N_{DC} , N_{RO} and N_Q respectively. From Fig. 1, neglecting the interference of CFA sampling and interpolation, the observed image intensity can be expressed as:

$$I = f(R) + N_Q = f(R_0 + N_{PS} + N_{DC} + N_{RO}) + N_Q \quad (2)$$

where $f(\cdot)$ is the CRF. By Taylor expansion, the first-order approximation is [7]:

$$I \approx f(R_0) + f'(R_0) \cdot (N_{PS} + N_{DC} + N_{RO}) + N_Q \quad (3)$$

Assume that all noise sources are zero-mean and independent of each other, and their standard deviations are σ_{PS} , σ_{DC} , σ_{RO} and σ_Q respectively. The variance of R and I can be written as:

$$\sigma_R^2 = \sigma_{PS}^2 + \sigma_{DC}^2 + \sigma_{RO}^2 \quad (4)$$

$$\sigma_I^2 \approx f'^2(R_0) \cdot (\sigma_{PS}^2 + \sigma_{DC}^2 + \sigma_{RO}^2) + \sigma_Q^2 \quad (5)$$

From (5), the variance of pixel values depends on 6 items: the shape of CRF, the value of noise-free radiance and the variances of each noise components. Photon shot noise is caused by quantum fluctuation and once superimposed by this type of noise, radiance satisfies the Poisson distribution [5], viz., σ_{PS} is proportional to the square root of R_0 . Both dark current noise and read-out noise can be treated as additive white noise in most scenes, unless the photo is taken in extreme dark with long exposures. Quantization noise is usually very low in comparison to the other types of noise, and therefore will be ignored in the present work. Thus, (5) can be simplified as:

$$\sigma_I^2 \approx f'^2(R_0) \cdot (R_0 \sigma_1^2 + \sigma_2^2) \quad (6)$$

where $\sigma_1^2 = \sigma_{PS}^2 / R_0$ and $\sigma_2^2 = \sigma_{DC}^2 + \sigma_{RO}^2$. Since NLF is defined as a function of I_0 , the noise-free image intensity, according to $R_0 = g(I_0)$, Equation (6) can be further rewritten as:

$$\sigma_I(I_0) \approx \sqrt{g(I_0) \sigma_1^2 + \sigma_2^2} / g'(I_0) \quad (7)$$

Now we have revealed the relationship between NLF σ_1 and inverse CRF g . Using (7), NLF can be found from g , σ_1 and σ_2 . The inverse CRF g depends on the camera model, and σ_1 and σ_2 are related to the image recording parameters such as ISO, shutter speed, aperture, and the camera specifications, e.g., the level of sensor noise. We model NLF based on (7) for the following two

considerations: 1) unlike the training based models [3], the parameters used here have physical meanings, and 2) estimation of NLF can be well constrained by the shape of CRF, especially in the relatively low noise level conditions. Therefore NLF and CRF can be well determined simultaneously.

3 NLF estimation

To estimate the noise level function from a single color image, a Bayesian approach is introduced. We divide the image into non-edge and edge regions, collect sample sets from both regions, and optimize the NLF using a maximum *a posteriori* (MAP) inference. Fig. 2 is a block diagram of the method, which consists of two parts: sample sets collection and the formulation of the likelihood model and Bayesian MAP framework. These will be discussed in the following section.

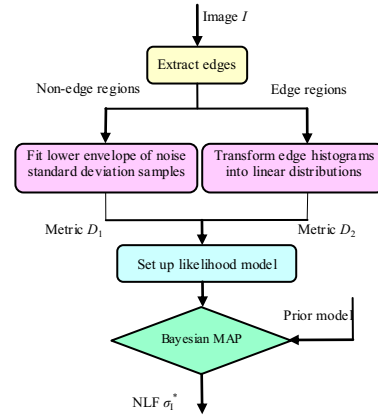


Figure 2. Block diagram of the proposed NLF estimation method.

3.1 Estimation sample sets collection and distance metrics definition

We first use the Canny operator and three-pixel dilation to divide the image into two groups of regions: edge regions and non-edge regions. Assume these regions are independent of each other. We group the non-edge pixels into non-overlapping regions based on spatial and intensity similarities using a mean-shift segmentation method as described in [8]. Let the original image be I , we apply a de-noise filter, denoted $d(\cdot)$, to get its smoothed version I_D . Here, $d(\cdot)$ is a wavelet based adaptive filter, which has been proved to be effective for images contaminated by common noise [9]. We define noise residual N as the difference between I and I_D . Then I_D and N are grouped into segments simultaneously according to mean-shift segmentation coordinates. Let I_i , I_{D_i} and N_i be the i -th segments in I , I_D and N , and \hat{I}_i and $\hat{\sigma}_i$ the average of I_{D_i} and standard deviation of N_i respectively. For each segment, we get a sample pair consisting of \hat{I}_i and $\hat{\sigma}_i$. If a segment N_i is small enough or $\hat{\sigma}_i$ is larger than the global standard deviation of N , we consider it as improper and discard it. With all proper

segments, we can get a sample set $\{(\hat{I}_i, \hat{\sigma}_i)\}$ for each color channel, *i.e.*, red, green and blue. However, since no ideal filter can separate image content and noise completely, we cannot estimate the noise level from an image directly from $\{(\hat{I}_i, \hat{\sigma}_i)\}$. There are two types of errors existing in N : the one is that some details of the image content are wrongly treated as noise component and left in N , and the other is that some real noise are regarded as image details and filtered out. Generally, probability of the former error type is larger than the latter. For this reason, we use the sum of lower envelope squared error as a measure of the difference between ground-truth NLF and collected samples. For each channel, to obtain the lower envelope of $\{(\hat{I}_i, \hat{\sigma}_i)\}$, we discretize the range of intensity $[0, 1]$ into uniform intervals $\{[nh, (n+1)h]\}$ with $n=0,1,\dots,(1/h-1)$, where h is a uniform interval and is set to $1/64$ in our method. For any subset $\Lambda_n = \{(\hat{I}_i, \hat{\sigma}_i) | nh \leq \hat{I}_i \leq (n+1)h\}$, we can find a pair $(\hat{I}_n, \hat{\sigma}_n)$ with the minimum $\hat{\sigma}_n = \min_{\Lambda_n} \hat{\sigma}_i$. By traversing all the possible Λ_n , we can get a sample set $\{(\hat{I}_n, \hat{\sigma}_n) | 0 \leq n \leq 1/h-1\}$ from each channel. Then we combine all sets collected from each channel as a whole set, denoted Ω_1 . Suppose the ground truth NLF is known, denoted σ_1 . Then we use the sum of squared error to measure the difference between σ_1 and samples $\hat{\sigma}_n$ in Ω_1 . Denote this difference metric as D_1 , which can be computed as

$$D_1(\sigma_1; \Omega_1) = \sum_{k=RGB} \sum_{n=0}^{1/h-1} [\sigma_{1k}(\hat{I}_{nk}) - \hat{\sigma}_{nk}]^2 \quad (8)$$

where subscript k is added for distinguishing all components in different color channels. Following the model derived from Equation (7), Equation (8) can be rewritten as

$$D_1(g, \sigma_1, \sigma_2; \Omega_1) = \sum_{k=RGB} \sum_{n=0}^{1/h-1} \left[\frac{\sqrt{g(\hat{I}_{nk})\sigma_{1k}^2 + \sigma_{2k}^2}}{g'(\hat{I}_{nk})} - \hat{\sigma}_{nk} \right]^2 \quad (9)$$

As indicated in Section 2, NLF is strongly dependent on the shape of camera response function. In our method, to restrict the profile of NLF, we employ an edge based inverse CRF estimation method proposed in [10]. Suppose a small patch in image containing two regions with distinct but uniform colors, denoted R_{1m}, R_{2m} . Let \hat{M}_{1m} and \hat{M}_{2m} denote the mean color of R_{1m} and R_{2m} respectively. In addition, assume a pixel \hat{M}_{Em} is on the edge between R_{1m} and R_{2m} . Here $\hat{M}_{1m}, \hat{M}_{2m}$ and \hat{M}_{Em} are all three-element vectors which represent intensities in RGB channels. As stated in [10], the irradiance of the pixel on edge should be a linear combination of those pixels on R_{1m} and R_{2m} before nonlinear CRF mapping. This property is used to estimate inverse CRF, namely finding a function g to map $\hat{M}_{1m}, \hat{M}_{2m}$ and \hat{M}_{Em} back to the linear relationship in color spaces. The distance from $g(\hat{M}_{Em})$ to line $\overline{g(\hat{M}_{1m})g(\hat{M}_{2m})}$ can be computed as:

$$H_m = \frac{[g(\hat{M}_{1m}) - g(\hat{M}_{2m})] \times [g(\hat{M}_{Em}) - g(\hat{M}_{2m})]}{|g(\hat{M}_{1m}) - g(\hat{M}_{2m})|} \quad (10)$$

where $|\cdot|$ is the Euclidean norm (L^2 -norm) of a vector and \times denote the cross product of two vectors. After scanning all patches along extracted edges, we select the patches containing two regions with distinct but uniform colors as valid sample patches for estimation. Suppose the total number of valid patches is K , then we get a sample set $\{(\hat{M}_{1m}, \hat{M}_{2m}, \hat{M}_{Em}) | 1 \leq m \leq K\}$, denoted Ω_2 . Given an inverse CRF g , we define a total distance metric D_2 to measure the linearity mapping by g :

$$D_2(g; \Omega_2) = \sum_{\Omega_2} H_m^2 \quad (11)$$

Here, different from [10], we define D_2 as the sum of squares instead of direct accumulation to achieve the consistency with the definition in D_1 . Although there are other existing algorithms for CRF estimation, we apply edge based method due to the following consideration: image edge regions and non-edge regions can be treated as independent components and this merit can greatly facilitate us to construct the likelihood model in the next step.

So far we have interpreted how to collect two independent sample sets Ω_1 and Ω_2 from image non-edge regions and edge regions. In addition, two distance metrics D_1 and D_2 have also been defined as the measurements of differences between estimate values and samples.

3.2 Estimating NLF using Bayesian MAP inference

To estimate the NLF from inadequate sample sets Ω_1 and Ω_2 , we use Bayesian MAP inference to solve this problem. Prior and likelihood models should be set up before maximizing the posterior probability.

Suppose g, σ_1 and σ_2 are independent with each other. Based on the model proposed in Equation (7), the prior model of NLF can be disassembled as:

$$P(\sigma_1) = P(g, \sigma_1, \sigma_2) = P(g) P(\sigma_1) P(\sigma_2) \quad (12)$$

where $P(g)$ is the prior probability of inverse CRF and $P(\sigma_1), P(\sigma_2)$ represent the prior probabilities of σ_1, σ_2 which are supposed to obey the uniform distribution. Due to the incomplete samples, it is impossible to estimate g with numerous parameters. To use fewer coefficients representing g , as indicated in [11], any g can be concisely represented as

$$g = \mathbf{g}_0 + \sum_{j=1}^J \alpha_j \mathbf{g}_j \quad (13)$$

where \mathbf{g}_0 and \mathbf{g}_j ($j=1,2,\dots,J$) represent the mean vector and eigenvectors of 201 real-world inverse CRF curves, respectively. All these vectors can be downloaded from the database of real-world CRFs (DoRF), which are derived from a principal component analysis (PCA) on the observed CRF data. α_j ($j=1,2,\dots,J$) are the

representation coefficients, namely, we can use limited (the number of J) coefficients α_j describing g . Note the lengths of vectors \mathbf{g}_0 and \mathbf{g}_j are all 1024 in this model. Following [10], J is set to 5 and the prior probability of g is formulated by a Gaussian mixture model which can be trained from the PCA coefficients in the DoRF database.

Likelihood function represents the probability to see observed sample sets once the parameters for estimation are given. Since Ω_1 and Ω_2 can be regarded as independent sets and the likelihood probability monotonically decrease with the increment of D_1 and D_2 , the likelihood function $L(\sigma_1)$ is defined as a joint bivariate exponential distribution:

$$\begin{aligned} L(\sigma_1) &= P(\Omega_1, \Omega_2 | g, \sigma_1, \sigma_2) \\ &= P(\Omega_1 | g, \sigma_1, \sigma_2) P(\Omega_2 | g) \\ &\propto \exp[-\lambda_1 D_1(g, \sigma_1, \sigma_2; \Omega_1) - \lambda_2 D_2(g; \Omega_2)] \end{aligned} \quad (14)$$

where likelihood L is proportional to a product of two negative exponential functions with respect to D_1 and D_2 , respectively. λ_1 and λ_2 are two weighting parameters to control the constraint intensity on the shape of NLF and are empirically both set to 10 in our method.

Now, based on the model of prior and likelihood, we solve the whole problem by Bayesian MAP inference. In General, for a color image, there are three different NLF curves in RGB channels individually. However, considering the similarity of g in each channel and ease of our estimation, the inverse CRF g in every channel is considered to be identical. Based on this assumption, we estimate all three NLFs totally using 11 parameters: $\{(\alpha_j, \sigma_{1k}, \sigma_{2k})\}$, where $j=1\dots 5$, $k=R, G, B$. Taking the logarithm of posterior function, Bayesian inference turns out to be a minimum problem:

$$\sigma_1^* = \arg \min_{\{(\alpha_j, \sigma_{1k}, \sigma_{2k})\}} [\lambda_1 D_1(g, \sigma_1, \sigma_2; \Omega_1) + \lambda_2 D_2(g; \Omega_2) - \log(P(g))] \quad (15)$$

To obtain the minimum of Equation (15), we seek the local minimum by non-linear Levenberg-Marquardt-Fletcher algorithm with 40 groups of different initial values. To reduce complexity, the maximum number of iterations for each group is limited to 200. Finally NLF is determined by selecting one group of optimal values from all candidate minimums followed by a low-pass filter to refine the shape of NLF.

4 Experimental results

To evaluate our algorithm, we compare NLFs from estimation with those of ground-truth. Following [3], the reference ground-true NLF is obtained by taking 30 photos of a static scene using the same camera with the same parameters and calculating the mean image. We collect 21 group test photos as our dataset shown in Fig.3. All photos are numbered consecutively from top left to bottom right, where 01-10 are captured by Nikon D300, 11-16 Nikon D3100, 17-19 Canon EOS 400D and 20-21 Sony A350. Each photo is saved by RAW, TIF or fine JPEG formats and cropped or down-sampled to a size of 1024×1024 square image for reduction of computational complexity. Fig.4 exemplifies our estimation results with

different camera models. The first column in Fig.4 depicts the test images, the rest columns show our estimated NLFs (labeled with color solid lines), observed sample sets Ω_1 (labeled with color dots) and reference true NLFs (labeled with gray dashed lines) in red, green and blue components, respectively. As is shown in Fig. 4, the agreements between our estimated NLFs and references, and the fits from the profile of observed samples are both performed well for most test images except for red component of test image no.14 listed in the second row of Fig.4. One cause of this deviation is due to the complicated texture in red component of test image no.14, and the other is the noise model in Nikon D3100 is more like an AWGN model than our supposed model.

For comparing with the method proposed in [3], we also use root-mean-square error (RMSE) and infinity norm (L^∞ -norm) to measure the difference between estimated NLFs and the reference NLFs. Here, L^∞ -norm of any vector can be simply determined by the absolute maximum of each element. Fig. 5 presents the RMSE of all test images in each color component. Table 1 lists the error statistics comparison of our method with Liu et al.'s method, where the means and variances under two metrics are counted. Observed from Fig. 5 and Table 1, the proposed method has a smaller statistical error between estimators and ground truths. It should be explained that statistics offered in [3] are measured from relatively higher synthetic noisy images, so in the final analysis, the main difference between the proposed method and [3] comes from different objectives. The proposed method is more suitable for real taken photos.

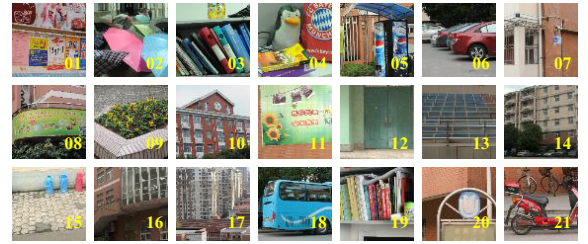


Figure 3. Test scenes for noise level function evaluation.

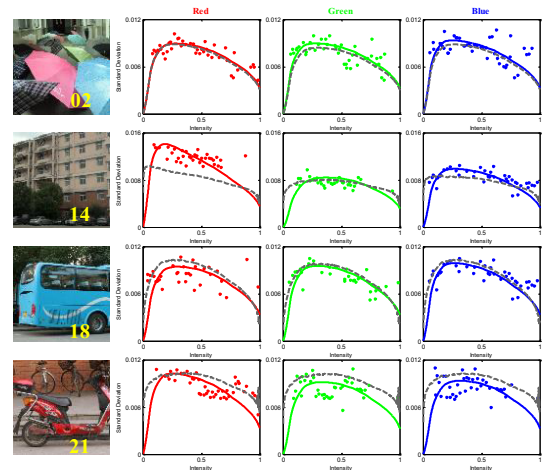


Figure 4. Examples of estimation of NLF from single image and comparison with 'ground-truth' NLFs extracted from multi-shot images.

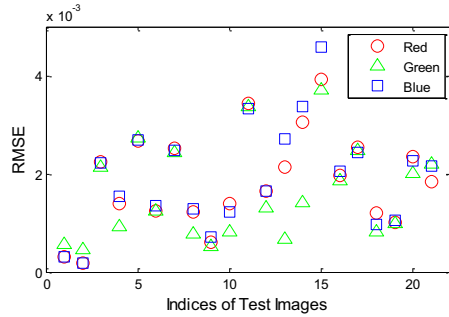


Figure 5. Root-mean-square error (RMSE) of each test

Table 1. The error statistics of the RMSE and L^∞ -norm between the estimated NLFs and the ground truths.

Error metrics	Proposed		Liu et al. [3]	
	mean	variance	mean	variance
RMSE	0.0018	9.54e-007	0.0048	1.09e-005
L^∞ -norm	0.0050	6.96e-006	0.0110	1.44e-004

5 Conclusions

In this paper, we have developed an algorithm to estimate the noise level function (NLF) from a single color image. Instead of a simple additive Gaussian noise assumption, we suppose the noise in most real photos is irradiance-dependent. Following the definition of noise level function, we collect noise statistical samples from every image segment. Due to the incomplete sample set, we develop an approximate model to represent NLF by a first-order Taylor series expansion. Deduced from this model, the profile of a NLF is strongly dependent on its corresponding camera response function (CRF). To constrain the shape of NLF, we employ an edge based CRF estimation method and integrate it into our estimation method by a Bayesian framework.

Experimental results have demonstrated the efficacy of our NLF estimation and forgery detection methods.

Acknowledgments

This work was supported in part by the Hujiang Foundation of China (B14002/D14002)

References

1. D. Zoran, Y. Weiss, Scale invariance and noise in nature image, Proc. 12th IEEE Int. Conf. Computer Vision, pp. 2209-2216 (2009)
2. B. Mahdian, S. Saic, Image and Vision Computing 27, 10, pp.1497-1503 (2009)
3. C. Liu, W. T. Freeman, R. Szeliski, S. B. Kang, Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition, 1, pp. 901-908 (2006)
4. C. Liu, R. Szeliski, S. B. Kang, C. L. Zitnick, W. T. Freeman, IEEE Trans. Pattern Anal. Machine Intel. 30, 2, pp. 299-314 (2008)
5. Y. Tsin, V. Ramesh, T. Kanade, Proc. 8th IEEE Int. Conf. Computer Vision, 1, pp. 480-487 (2008)
6. G. E. Healey, R. Kondepudy, IEEE Trans. Pattern Anal. Machine Intel. 16, 4, pp. 267-276 (1994)
7. J. Takamatsu, Y. Matsushita, K. Ikeuchi, Proc. 10th European Conf. Computer Vision, 4, pp.623-637 (2008)
8. D. Comaniciu, P. Meer, IEEE Trans. Pattern Anal. Machine Intel. 24, 5, pp. 603-619 (2002)
9. J. Lukas, J. Fridrich, M. Goljan, IEEE Trans. Inf. Forensics Security 1, 2, pp. 205-214(2006)
10. S. Lin, J. Gu, S. Yamazaki, H. Y. Shum, Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition 2, pp. 938-945 (2004)
11. M.D. Grossberg, S.K. Nayar, IEEE Trans. Pattern Anal. Machine Intel. 26, 10, pp. 1272-1282 (2004)