SOURCE CAMERA IDENTIFICATION BASED ON CAMERA GAIN HISTOGRAM

*Sz-Han Chen and Chiou-Ting Hsu**

Department of Computer Science, National Tsing Hua University, Taiwan
Email: cthsu@cs.nthu.edu.tw, Phone: +886-3-5742960, Fax: +886-3-5723694

ABSTRACT

In this paper, we propose a novel approach for source camera identification based on camera gain histogram. By using the photon transfer curve (PTC) as camera noise model, we construct camera gain histogram from the occurrences of different camera gain constants. With the distribution of camera gain histogram for each camera, we extract four features to characterize the camera. In our experiments, 400 photos acquired from two high-end digital cameras at two different exposure levels are used to evaluate the effectiveness of the proposed approach. A two-class support vector machine (SVM) is employed as a classifier. Our experimental results demonstrate that the distinction rate in identifying different cameras achieves promising performance.

Index Terms — source camera identification, CCD noise, photon transfer curve, camera gain histogram

1. INTRODUCTION

With the advances in digital image processing techniques, digital photos can now be easily manipulated, synthesized and tampered in numerous ways without leaving visible clues. Even though when nowadays photos are largely used as important evidence in court, “authenticity” of digital photos is still a big question mark. As a consequence, source camera identification [1, 2, 3] becomes a hot topic to help checking the trustiness of digital photos.

The goal of source camera identification is to determine which camera is used to capture the test photos. The basic idea of source camera identification comes from the manufacturing process of cameras. Although most digital cameras can store the imaging-related information in the output file header, such as camera manufactory, camera model, exposure level, and date of photos, the contents in the file header could be easily modified or removed using image editing tools. Therefore the photo file header can no longer provide reliable information for identifying source camera.

As CCD sensor is not a perfect device in which the sensor output carries not only pure signal but also various noise components, sensor noise model could be used as a representative feature for cameras. For example, at low illumination level, dark current noise which generated in totally dark condition dominates the sensor noise. As the illumination increases, the sensor noise becomes more signal-dependent. Therefore, photon transfer curve (PTC) is useful to represent the relationship between the sensor output and the corresponding noise [4]. By plotting the noise standard deviation as a function of the sensor output on a log-log scale, we can use PTC to characterize camera noise. For example, using PTC as the camera noise model for the development of noise filtering has been proposed in [5, 6].

This paper aims to discuss the source camera identification problem based on the distribution of camera gain histogram. In a typical CCD imaging process, camera gain constant is a very important conversion constant. It is viewed as a combined parameter accomplished by a series of functional blocks in a camera. Once the PTC is estimated, the camera gain constant is computed by the sensor output value and the corresponding noise variance. By recording the occurrences of different camera gain constant values, we can then generate the camera gain histogram. Next, we proposed to extract four features from the camera gain histogram to characterize the camera noise model. Finally, we adopt the support vector machine (SVM) for the classification to distinguish different source cameras.

The rest of this paper is organized as follows. Section 2 explains the proposed approach and the formulations. Section 3 shows the experimental results. Finally, Section 4 gives the conclusion.

2. PROPOSED APPROACH

2.1. Camera noise model

In this section, we first describe the relationship between the photon transfer and the camera noise model. The photon transfer technique is certified to be helpful for calibrating and characterizing the CCD systems [4]. A photon transfer curve (PTC) is plotted under various camera sensor output and corresponding noise standard deviation on a log-log scale. Since camera RAW output is analogous to the camera sensor output; PTC can be approximated by the noise curve from RAW photos. An ideal PTC is composed of three distinct regions, according to the slope changes (as shown in Fig. 1). The first region is characterized by a flat curve at the lower illumination level. In this region, the intensity of

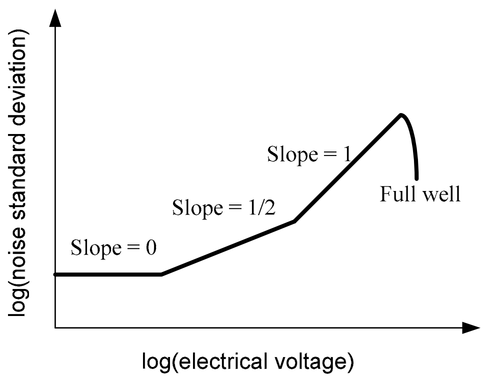


Fig. 1 An ideal photon transfer curve (PTC) of a camera.

noise is independent of the sensor output, such as dark current noise and read-out noise. As the CCD system receives much more photons, the noise tends to be signal-dependent. The curve with slope of approximately 1/2 depicts the second region. Within this region, the shot noise is dominant at the middle illumination level. The third region is denoted at higher illumination level, indicating that the intensity of noise is proportional to the sensor output; hence the slope of curve in this region is 1. Although the fixed pattern noise is dominated in this region, it can be removed by modern digital camera using flat fielding. The suddenly decrease of the noise intensity at higher illumination level is due to the physical full well. Hence, we can feasibly assume a two-region PTC by eliminating the affect of the third region in our experiment.

2.2. Camera gain histogram

Camera gain histogram is a graphical version designed to display occurrences of camera gain constant. The camera constant K is defined as

$$K = \frac{\mu}{\sigma_s^2 - \sigma_d^2}, \tag{1}$$

where μ is the average pixel value, σ_s^2 and σ_d^2 are the variance corresponding to the signal-dependent and signal-independent noise components. Although the value of camera gain constant K is affected by various μ, σ_s^2 and σ_d^2, we can still obtain a camera gain histogram using the summarized information of different K and its occurrences. Fig. 2 shows two instances of simulated camera gain histogram for two digital cameras used in our experiment.

2.3. Feature extraction

Given the distribution of camera gain histogram for each camera, we proposed to extract four statistical features from the histogram to characterize the camera. The four features we adopted in this paper are as follows:

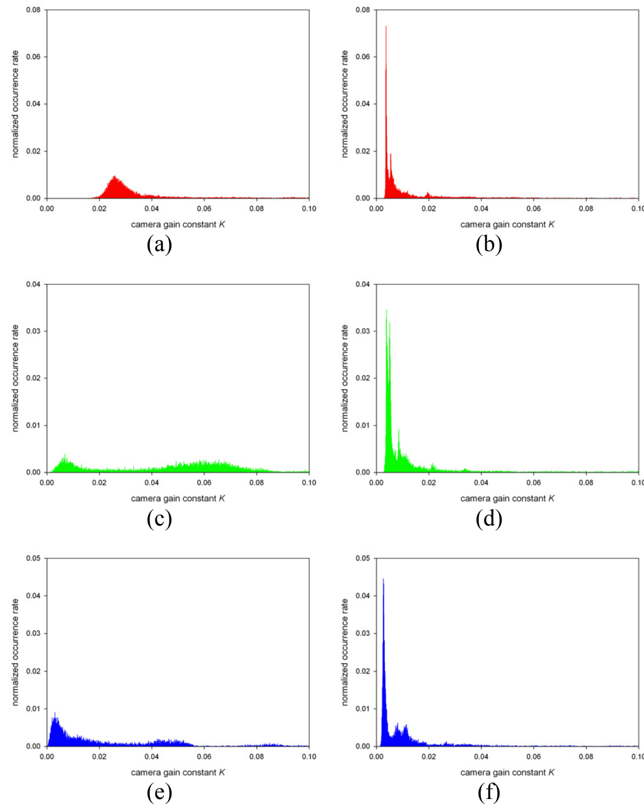


Fig. 2 The camera gain histogram per color channel (RGB) of two digital cameras: (a), (c) and (e) belong to Nikon D80; (b), (d) and (f) belong to Canon EOS 350D. The occurrences are normalized in order to emphasize the difference between the distributions of two camera gain histograms.

1) Mean (F_m)

$$F_m = \sum_{h=0}^{L} hP(h), \tag{2}$$

2) Standard Deviation (F_d)

$$F_d = \sum_{h=0}^{L} (h - F_m)^2 P(h), \tag{3}$$

3) Energy (F_e)

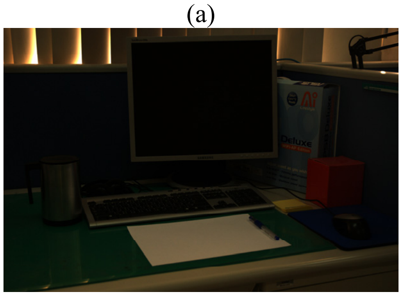
$$F_e = \sum_{h=0}^{L} (P(h))^2 \text{, and} \tag{4}$$

4) Entropy (F_t)

$$F_t = -\sum_{h=0}^{L} P(h)\log_2(P(h)), \tag{5}$$

where h denotes the quantized value of camera gain

(a)



(b)

Fig. 3 (a) The photo captured at exposure level 1 and (b) the photo captured at exposure level 2.

Table 1 Parameter Settings of two cameras

	Exposure level 1	Exposure level 2	ISO	Maximal pixels
Nikon D80	F 4.5 / 80	F 4.5 / 125	800	3872×2592
Canon EOS 350D	F 4.5 / 80	F 4.5 / 125	800	3456×2304

constants for $0 \leqq h \leqq L$. The estimation of $P(h)$ is simply defined by

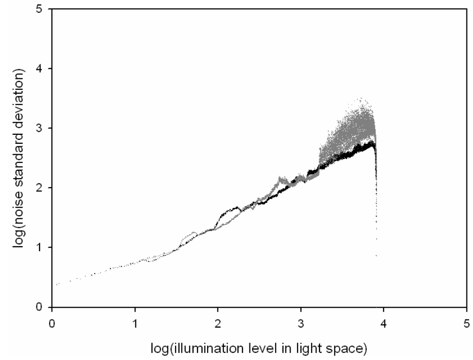
$$P(h) = \frac{N(h)}{M}, \tag{6}$$

where $N(h)$ indicates the number of occurrences of camera gain constant corresponding to the value h, and M is the total number of occurrences in the camera gain histogram. We then use a four dimensional feature vector $\mathbf{F}$ to represent the distribution of camera gain histogram

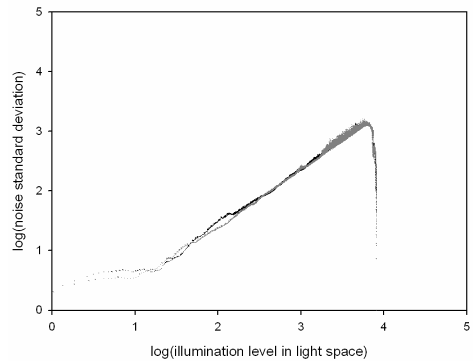
$$\mathbf{F} = (F_m, F_d, F_e, F_t)^T. \tag{7}$$

2.4. Classification

In this paper, we use a two-class SVM as the classifier to classify two different sources of digital cameras. The four-dimensional feature vectors belong to each digital camera constitute a feature set. This feature set is further divided into a training set and a test set for SVM [4] training and performance evaluation, respectively. Gaussian kernel is utilized in our experiment as the kernel function of the SVM classifier.



(a)



(b)

Fig. 4 Estimated PTC at different exposure level of (a) Nikon D80 and (b) Canon EOS 350D. The noise curve of exposure level 1 is represented by black color, and the noise curve of exposure level 2 is represented by gray color.

3. EXPERIMENTS

3.1. Photo data set

The photos used in our experiment are acquired from two high-end digital cameras: Nikon D80 and Canon EOS 350D. Both cameras are set to output the 12-bit RAW image format with the maximal pixels and the highest quality. The parameter settings of cameras are shown in Table 1.

At two different exposure levels, we use high ISO value of 800 to generate sufficient noisy photos. In order to exclude the influences from unpredictable variation (such as winds, clouds or lighting changes), outdoor scenes are currently not adopted in our experiment. The photos we used in the experiments mainly contain indoor scenes. One example is shown in Fig. 3. We use two cameras to capture 400 photos for our experiment, in which 100 scenes are captured at two different exposure levels for each camera.

In addition, we also acquire 50 photos without any irradiance from each camera. We use these photos to estimate the noise standard deviation (σ_d) at lower illumination level for the two cameras.

Table 2. Identification results at exposure level 1 between two cameras

		Predicted		Accuracy (%)
		Nikon	Canon	
Actual	Nikon	21	0	100%
	Canon	0	21	100%

Table 3. Identification results at exposure level 2 between two cameras

		Predicted		Accuracy (%)
		Nikon	Canon	
Actual	Nikon	17	4	81%
	Canon	2	19	90%

Table 4. Identification results between two cameras using the combined exposure level

		Predicted		Accuracy (%)
		Nikon	Canon	
Actual	Nikon	36	6	86%
	Canon	4	38	90%

3.2. PTC estimation and comparison

In order to estimate the PTC from light space, we use a series of 100 RAW photos at a fixed exposure level acquired from the camera. All the noise standard deviations are computed as the standard deviations corresponding to each intensity value in the averaged image over all photos. The estimation of PTCs are shown in Fig. 4. As Fig. 4 shows, two cameras produce similar PTCs, while the Nikon D80 generates much more fluctuant curves than Canon EOS 350D does.

3.3. Camera gain histogram and feature extraction

From the estimated PTC, we can obtain numerous values of intensity (μ), standard deviation (σ_s) of signal-dependent noise and standard deviation (σ_d) of signal-independent noise. Substituting μ, σ_s and σ_d into Eq. (1), we then obtain the camera gain constant K. Next, we construct the camera gain histogram by calculating the occurrences of different K value in a quantized manner. A four dimensional feature vector is extracted from the camera gain histogram including mean, standard deviation, energy and entropy. After the feature extraction, the SVM classifier is utilized for training and testing.

3.4. Results

In the first part, we investigate the source camera identification at different exposure level. The photos at a fixed exposure level are divided into 20 sets, and every set contains 5 photos. In each set, we compute the camera gain histogram and perform the feature extraction per color channel. Thus we obtain 60 samples at the exposure level, where 39 samples are chosen as training samples and the other 21 samples are the test samples. The identification results for exposure level 1 and 2 are shown in Table 2 and Table 3, respectively. The identification rate is much higher at the exposure level 1 than at the exposure level 2. One possible explanation is that the noise pattern is unobservable at the lower exposure level; hence the classification performance is degraded.

In the second part, we combine the photo data sets from the two different exposure level of the same camera. The photos belong to one of the two camera are divided into 40 sets, and every set contains 5 photos. Similarly, in each set, we compute the camera gain histogram and extract features per color channel. Here, we obtain 120 samples for a camera. We choose 78 samples as the training samples and the other 42 samples as the test samples. The identification results are shown in Table 4. From the results of Table 4, we observe that the identification rate of Nikon D80 is slightly worse than Canon EOS 350D. It is not surprised to obtain this result since the noise curve of Nikon D80 is unstable as shown in Fig. 4.

4. CONCLUSION

In this paper, we propose a novel approach for the identification of source cameras. The PTC is first estimated as the noise curve using RAW photos, and the camera gain histogram is then derived from the occurrences of different camera gain constants. With the camera gain histogram, we extract four features from the distribution and then use the features for training and testing the SVM classifier. The photos acquired from two cameras at different exposure levels are adopted as our experiment data sets. Experimental results show that the identification rate is satisfactory while distinguishing two cameras from different manufactories. Nevertheless, we also find that the performance is poorer at lower exposure level. Further investigation on the camera noise pattern will be our major future work.

5. REFERENCE

[1] J. Lukãš an J. Fridrich, "Digital camera identification from sensor pattern noise," *IEEE Trans. Information Forensics and Security*, vol. 1, no. 2, pp. 205-214, Jun., 2006.

[2] M. Kharrazi, H. T. Sencar and N. Memon, "Blind source camera identification," *Proc. ICIP*, 2004.

[3] S. Bayram, H. Sencar, N. Memon and I. Avcibas, "Source camera identification based on CFA interpolation," *Proc. ICIP*, 2005.

[4] J. R. Janesick, *Scientific Charged-Coupled Devices*. Bellingham, WA: SPIE 2001.

[5] H. Faraji and W. J. MacLean, "CCD noise removal in digital images," *IEEE Trans. Image Processing*, vol. 15, no. 9, pp. 2676-2685, Sep., 2006.

[6] H. Faraji and W. J. MacLean, "Adaptive suppression of CCD signal-dependent noise in light space," *Proc. ICASSP*, 2004.

[7] C. C. Chang and C. J. Lin, "LIBSVM: a library for support vector machines," 2001. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm