Building Fingerprints with Information from Three Color Bands for Source Camera Identification

Yongjian Hu
School of Electronic and Information Engineering, South China University of Technology
Guangzhou 510641, China
E-mail: eeyjhu@scut.edu.cn

Chang-Tsun Li
Department of Computer Science, University of Warwick
Coventry CV4 7AL, UK
E-mail: c-t.li@warwick.ac.uk

Chao Jian
College of Automation Science and Engineering, South China University of Technology
Guangzhou 510641, China
E-mail: 79830849@qq.com

ABSTRACT
The imaging sensor (e.g., CCD, CMOS) pattern noise is a noise-like spread-spectrum signal that mainly consists of the photo-response non-uniformity (PRNU) noise. Since each camera has a unique PRNU, the PRNU becomes a popular fingerprint for source camera identification. Current fingerprints are built based on one color band. To better reflect the intrinsic characteristics of the camera, this work explores the way to build fingerprints based on three color bands of an image.

Categories and Subject Descriptors
I.4 [**Image Processing**]: Miscellaneous

General Terms
Algorithms, Security

Keywords
Digital Image Forensics, Camera Identification, Sensor Pattern Noise, Camera Fingerprint, Color Band

1. INTRODUCTION
The photo response non-uniformity noise of imaging sensors was first used by Lukas *et al.* as a device fingerprint for source camera identification [1]. Since then, the PRNU-based passive authentication methods have attracted lots of attention and have been used for camera/scanner identification (e.g., [1]-[10]) as well as digital image forgery detection (e.g., [3]).

The most important problem of PRNU-based identification methods is the extraction of the PRNU signal. Lukas *et al*. proposed several ways to extract the PRNU from an image in [1], such as using flat fielding, averaging images taken by the same camera, or directly using the average image of residual images. Due to the limitation of acquiring the raw sensor output and a large number of images in common situations, the first two options are relatively difficult to implement. The last option is more practical. A residual image can be obtained by subtracting a denoised version from the original image.

MiFor'10, October 29, 2010, Firenze, Italy.

The average image of dozens or even hundreds of residual images usually has a cleaner PRNU signal due to the counteraction of random noise components inherent in the original image (e.g., shot noise, read-out noise, and quantization noise). However, it still has some noise such as the shaping noise introduced by denoising filters. So Chen *et al*. [3] proposed a maximum-likelihood estimator to estimate a more accurate PRNU from the residual images. But the noise in the residual images is colored Gaussian noise rather than WGN (white Gaussian noise). Thus, the estimation of the PRNU has to be performed block-by-block across the image under the assumption that the noise is approximately WGN in a small image region. Machine learning was used to obtain the correlation predictor of each block. One obvious weakness of [3] is that the detection has to be taken block-by-block and the computational cost is high.

Due to the computational simplicity of the method in [1], some methods were proposed to improve it. For example, signal preprocessing operations were used to remove the scene content, stochastic noise and DCT blocking artifacts from the residual image in [4]. A signal enhancer was proposed to suppress the unwanted components of the residual image and enhance the components of interest in [5]. A way of reducing the effect of irrelevant noise sources on the average image was presented in [6].

Current PRNU-based camera identification methods make decisions relying upon the detection result from one color band (i.e., one RGB channel image) or relying upon pooling the results from three individual color bands. This is because their fingerprints are constructed based on individual color bands. We think that such fingerprints are difficult to comprehensively reflect the characteristics of the sensor array. In this work, we propose three schemes to build camera fingerprints with information from all the three color bands of an image. Although some methods (e.g., [7], [8], [11]-[13]) were proposed to use the information from three color bands for imaging device identification or forgery detection, they adopted statistical image features like the mean, variance, skewness and kurtosis. Such image features form a feature vector which is used to drive a pattern classifier like a support vector machine (SVM).

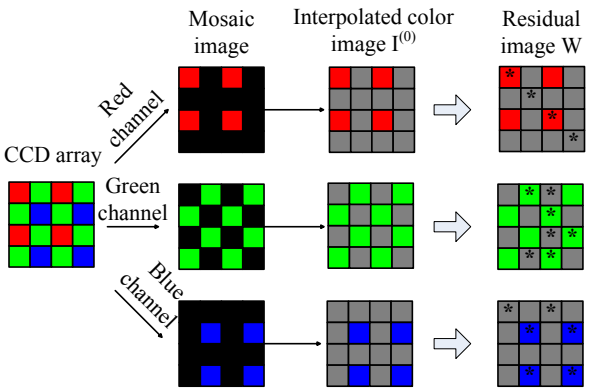

Figure 1. Obtaining the residual image. Black blocks refer to non-sampled values. Gray blocks refer to interpolated values. Blocks with * refer to the significant components.

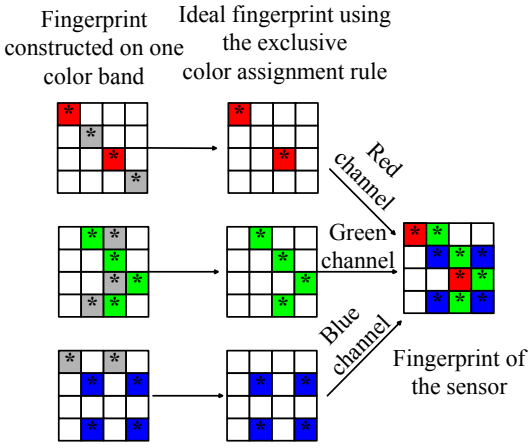

Figure 2. Construction of the camera fingerprint. Blank blocks represent zero elements.

2. BUILDING CAMERA FINGERPRINTS ON ONE COLOR BAND

In this section, we first introduce our scheme to build a camera fingerprint with information from one color band. We model an image I in the vector form of the sensor output. To simplify the notation and avoid introducing too many symbols, we directly adopt the expression given in [3]

$$I = I^{(0)} + I^{(0)}K + \Theta = (1+K)I^{(0)} + \Theta \qquad (1)$$

where $I^{(0)}$ is the sensor output in the absence of noise, K is the PRNU factor, and Θ is a complex of independent random noise components. Similar to [3], we assume that we have the camera or at least some images taken by the camera. We estimate K from the images taken by the camera. To improve the signal-to-noise ratio, it is desirable that $I^{(0)}$ be subtracted from (1) to emphasize the PRNU signal $I^{(0)}K$. In practice, however, we can only use the denoised version of the observed data I to approximate $I^{(0)}$. So the residual image W is

$$\begin{aligned} W &= I - F(I) \\ &= (1+K)\cdot(I^{(0)} - F(I^{(0)})) + (\Theta - F(\Theta)) \end{aligned} \qquad (2)$$

where F is the denoising filter. We further obtain the average image of n residual images as

$$\begin{aligned} P = \frac{1}{n}\sum_{l=1}^{n} W_l &= \frac{1}{n}\sum_{l=1}^{n}(1+K_l)\cdot(I_l^{(0)} - F(I_l^{(0)})) \\ &+ \frac{1}{n}\sum_{l=1}^{n}(\Theta_l - F(\Theta_l)) \end{aligned} \qquad (3)$$

Since the average image P is an approximation of the fixed pattern that remains in every image taken by the camera, the method in [1] simply used it as the camera fingerprint. The later method [3] tried to estimate K from W. From (2), however, it can be seen that both $I^{(0)} - F(I^{(0)})$ and $\Theta - F(\Theta)$ would affect the extraction of K. Θ comes from different noise sources, so it is not easy to completely remove the influence of $\Theta - F(\Theta)$. On the other hand, $I^{(0)}$ is basically an interpolated RGB channel image. Color interpolation algorithms often introduce interpolation artifacts. As a result, the information about the sensor array is not accurate at interpolated pixels. Moreover, F has some effect on $I^{(0)} - F(I^{(0)})$. Taking into account these factors, we propose a way to avoid estimating K from W. We choose from P the pixels with magnitude larger than a predefined threshold to construct a fingerprint, $\widetilde{P}$. The reason is that, if $|\frac{1}{n}\sum_{l=1}^{n}(1+K_l)\cdot(I_l^{(0)} - F(I_l^{(0)}))| >> |\frac{1}{n}\sum_{l=1}^{n}(\Theta_l - F(\Theta_l))|$ at a pixel, we have $P \approx \frac{1}{n}\sum_{l=1}^{n}(1+K_l)\cdot(I_l^{(0)} - F(I_l^{(0)}))$. Apparently, this pixel contains more reliable information of K. We will use these large-magnitude pixels for detection and ignore the pixels with magnitude below the predefined threshold. We call the selected pixels *significant components/elements*. $\widetilde{P}$ is the same size as P, but its elements are set to zero except those significant components. Usually, $\widetilde{P}$ is a sparse 2-D matrix. The process of building the camera fingerprint on one color band is illustrated in Figure 1. Note that simple signal preprocessing operations can increase the quality of the residual images [3], and thus improve our scheme.

3. BUILDING CAMERA FINGERPRINTS WITH INFORMATION FROM THREE COLOR BANDS

A CCD sensor is color blind. Thus, cameras equipped with a CFA (color filter array) only capture one color at each pixel, and the remaining colors must be interpolated from neighboring pixels. We call this *exclusive color assignment rule* taken by the CFA. In Section 2, we have explained how to build a camera fingerprint with information from one color band. Now we extend this scheme to three color bands.

Let subscripts r, g, b denote red, green, and blue colors, respectively. Our strategy is that we first apply the scheme stated in Section 2 to each individual color band to construct $\widetilde{P}_r$, $\widetilde{P}_g$ and $\widetilde{P}_b$, and then use the color relationship among the three RGB channels to combine $\widetilde{P}_r$, $\widetilde{P}_g$ and $\widetilde{P}_b$ into a new fingerprint. We describe an ideal situation before presenting our practical schemes.

$\widetilde{P}_r$, $\widetilde{P}_g$ and $\widetilde{P}_b$ may contain significant components that are not physically captured by the sensor. Interpolated samples often have interpolation artifacts that would affect the estimation of the PRNU values. If we know the CFA structure, we can easily remove the significant components obtained from the interpolated ones in $\widetilde{P}_r$, $\widetilde{P}_g$ and $\widetilde{P}_b$ and build a combined camera fingerprint, $\widetilde{P}_c$, as shown in Figure 2. In practical applications, however, we often do not have knowledge about the CFA structure. Thus it is not easy to implement this scheme.

In this work, three practical schemes are proposed. Let (i, j) refer to the coordinate of a pixel and $|\cdot|$ the absolute operation. Let $x, x_1, x_2 \in \{r, g, b\}, \quad x_1 \neq x_2, y \in \{x_1, x_2\}$. Our first scheme is based on the aforementioned analysis that a significant component of P often can better reflect the PRNU information than a small component. So Scheme 1 can be described as follows:

$$\widetilde{P}_c(i, j) = \widetilde{P}_k(i, j) \tag{4}$$

where $k = \arg\max |\widetilde{P}_x(i, j)|$. Apparently, no matter whether $\widetilde{P}_c(i, j)$ is physically captured by the sensor or interpolated, we always let it equal the largest one among $\widetilde{P}_r(i, j)$, $\widetilde{P}_g(i, j)$ and $\widetilde{P}_b(i, j)$.

To reduce non-zero components in $\widetilde{P}_c(i, j)$, we propose Scheme 2 as follows:

$$\widetilde{P}_c(i, j) = \begin{cases} \widetilde{P}_t(i, j), & \text{if } \widetilde{P}_y(i, j) \neq 0, \ \forall y \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

where $t = \arg\max |\widetilde{P}_y(i, j)|$. In this case, at least two of $\widetilde{P}_r(i, j)$, $\widetilde{P}_g(i, j)$ and $\widetilde{P}_b(i, j)$ should be non-zero components when constructing a non-zero component of $\widetilde{P}_c$.

To further reduce non-zero components in $\widetilde{P}_c$, Scheme 3 is given as follows:

$$\widetilde{P}_c(i, j) = \begin{cases} \widetilde{P}_k(i, j), & \text{if } \widetilde{P}_x(i, j) \neq 0, \ \forall x \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

where $k = \arg\max |\widetilde{P}_x(i, j)|$. In this case, all of $\widetilde{P}_r(i, j)$, $\widetilde{P}_g(i, j)$ and $\widetilde{P}_b(i, j)$ should be non-zero components when constructing a non-zero component of $\widetilde{P}_c$.

To better understand formulas (4), (5), and (6), we give an example of practical results in Figure 3. $\widetilde{P}_r$, $\widetilde{P}_g$ and $\widetilde{P}_b$ are shown on the upper row. Both gray and white blocks refer to the significant components selected using the scheme in Section 2. Specifically, white blocks within $\widetilde{P}_r$, $\widetilde{P}_g$ and $\widetilde{P}_b$ are the immediate results after using the exclusive color assignment rule. Shown on the lower row are the fingerprints constructed with Schemes 1, 2, and 3, respectively. It can be seen that, due to the stricter constraints, Scheme 3 creates much less non-zero elements than the other two schemes.

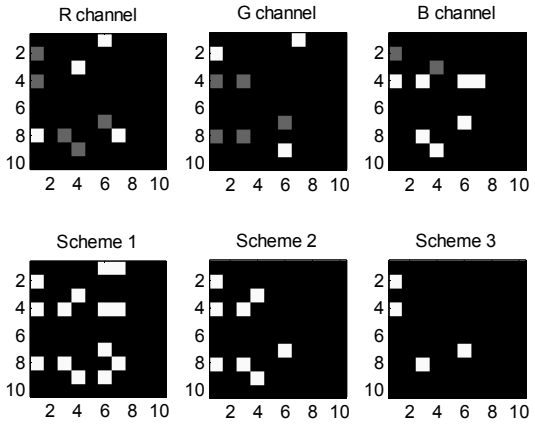

Figure 3. Practical construction example of the camera fingerprint. White blocks refer to the significant components selected using the joint information from the three color bands.

4. EXPERIMENT AND DISCUSSION

We have tested our schemes on different cameras, including Canon series A40, A620 and 450D, Nikon L3, Sony T10 and Olympus U820. Each camera takes 300 real-world photos saved in JPEG format at the highest resolution each camera can support, among which 150 photos are used for computing the reference camera fingerprint, and the rest for test. For a fair comparison, the image size is set to 1024×1024 no matter what a test camera is. Each image is acquired from a fixed part (e.g., the upper left corner) of a photo. All residual images are obtained using the denoising filter in [14], which has similar performance to what Lukas *et al*. used in [1] but runs faster. The method in [1] is performed on the green channel image. When applying our proposed method to each color band, we choose the top 5% of elements of P in magnitude as the significant components. The correlation detector employed in this work is the same as that in [1].

Camera identification can be formulated as a binary hypothesis testing problem. For the response of a correlation detector, the left peak, which indicates the absence of the reference camera fingerprint in test images, is always centered at 0 whereas the right peak, which indicates the presence of the reference camera fingerprint in test images, is centered at somewhere on the right side of the horizontal axis. The wider the two response peaks are, the more reliable the detector is. We can see from Figure 4 that, for the method proposed in [1], although the two response distribution modes in the plots associated with Canon A40 and 450D are separated, the overlapping modes in the other four plots

reveal the limitation of this method. On the other hand, from Figures 5 to 7, the well-separated distribution modes indicate that all our three schemes outperform the method proposed in [1]. We can also see that the right peak in all plots is moved rightward further and further from Figures 4 to 7, and the distance between the two peaks in all plots in Figure 7 is greater than that in Figure 4 to 6. This indicates that Scheme 3 is superior to the other three methods in the detection capability. Another merit of Scheme 3 is that it uses much fewer pixels for correlation detection than the method proposed in [1], which employs all the 1024×1024 pixels. Furthermore, Scheme 3 uses fewer pixels than Schemes 1 and 2. This also explains why the distributions for Scheme 3 are relatively flatter.

5. CONCLUSION

We have proposed a new way to build camera fingerprints for correlation-based camera identification. Instead of relying on information from one color band, the proposed three schemes use information from all the three color bands of an image. These strategies enable our camera fingerprints to better reflect the characteristics of the camera. Experimental results have proved the effectiveness of our three schemes for camera identification.

Signal preprocessing can suppress irrelevant noise components in the residual images while the knowledge of the CFA structure can help skip the interpolated pixels during the construction of fingerprints. Both will improve the quality of camera fingerprints. Our future work will make efforts in these directions.

6. ACKNOWLEDGMENTS

This work was supported by NSF of China 60772115.

7. REFERENCES

[1] Lukas J., Fridrich J., and Goljan, M. 2006. Digital camera identification from sensor pattern noise. *IEEE Transaction on Information Forensics and Security*. 1, 2 (Jun. 2006), 205-214.

[2] Fridrich J., Goljan M., and Filler, T. 2009. Large scale test of sensor fingerprint camera identification. In *Proceedings of SPIE, Electronic Imaging, Security and Forensics of Multimedia Contents XI* (San Jose, CA, Jan. 2009).

[3] Chen M., Fridrich J., Goljan M., and Lukas, J. 2008. Determining image origin and integrity using sensor noise. *IEEE Transaction on Information Forensics and Security.* 3, 1 (Mar. 2008), 74-90.

[4] Alles E. J., Geradts Z. J. M. H., and Veenman, Cor J. 2008. Source camera identification for low resolution heavily compressed images. In *Proceedings of International Conference on Computational Sciences and its Applications* (2008), 557-567.

[5] Li, C.-T. 2010. Source camera identification using enhanced sensor pattern noise. *IEEE Transaction on Information Forensics and Security.* 5, 2 (Jun. 2010), 280-287.

[6] Hu Y., Yu B., and Jian, C. 2009. Source camera identification using significant components of sensor pattern noise. In *Proceedings of International Conference on Computer Science and its Applications* (Korea, Dec. 10-12, 2009), 575-579.

[7] Filler T., Fridrich M., and Goljan, M. 2008. Using sensor pattern noise for camera model identification. In *Proceedings of IEEE International Conference on Image Processing* (2008), 1296-1299.

[8] Khanna N., Mikkilineni A.K., and Delp, E. J. 2009. Scanner identification using feature-based processing and analysis. *IEEE Transaction on Information Forensics and Security.* 4, 1 (Mar. 2009), 123-139.

[9] Caldelli R., Amerini I., and Picchioni, F. 2010. A DFT-based analysis to discern between camera and scanned images. *International Journal of Digital Crime and Forensics*. 2, 1 (Jan. - Mar. 2010), 21-29.

[10] Amerini I., Caldelli R., Cappellini V., Picchioni F., and Piva, A. 2010. Estimate of PRNU noise based on different noise models for source camera identification. *International Journal of Digital Crime and Forensics*. 2, 2 (Apr.-Jun. 2010), 21-33.

[11] Gallagher A. and Chen, T. 2008. Image authentication by detecting traces of demosaicing. In *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition* (Alaska, US, 2008), 1-8.

[12] Cao H. and Kot, A.C. 2009. Accurate detection of demosaicing regularity for digital image forensics. *IEEE Transaction on Information Forensics and Security.* 4, 4 (Dec. 2009), 899-910.

[13] Popescu A.C. and Farid, H. 2005. Exposing digital forgeries in color filter array interpolated images. *IEEE. Transactions on Signal Processing*. 53, 10 (Oct. 2005), 3948-3959.

[14] Chang G., Yu B., and Vetterli, M. 2000. Spatially adaptive wavelet thresholding with context modeling for image denoising. *IEEE Transaction on Image Processing*. 9, 9 (Sept. 2000), 1522-1531.

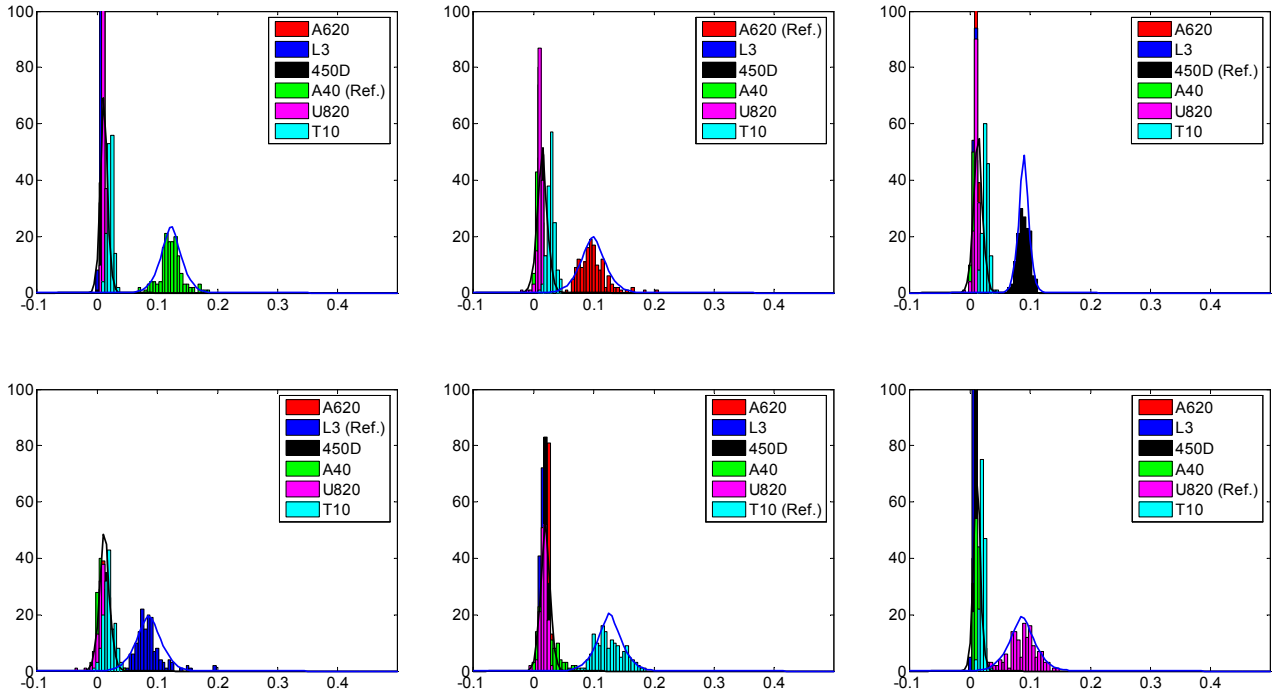

Figure 4. Results of the method proposed in [1].

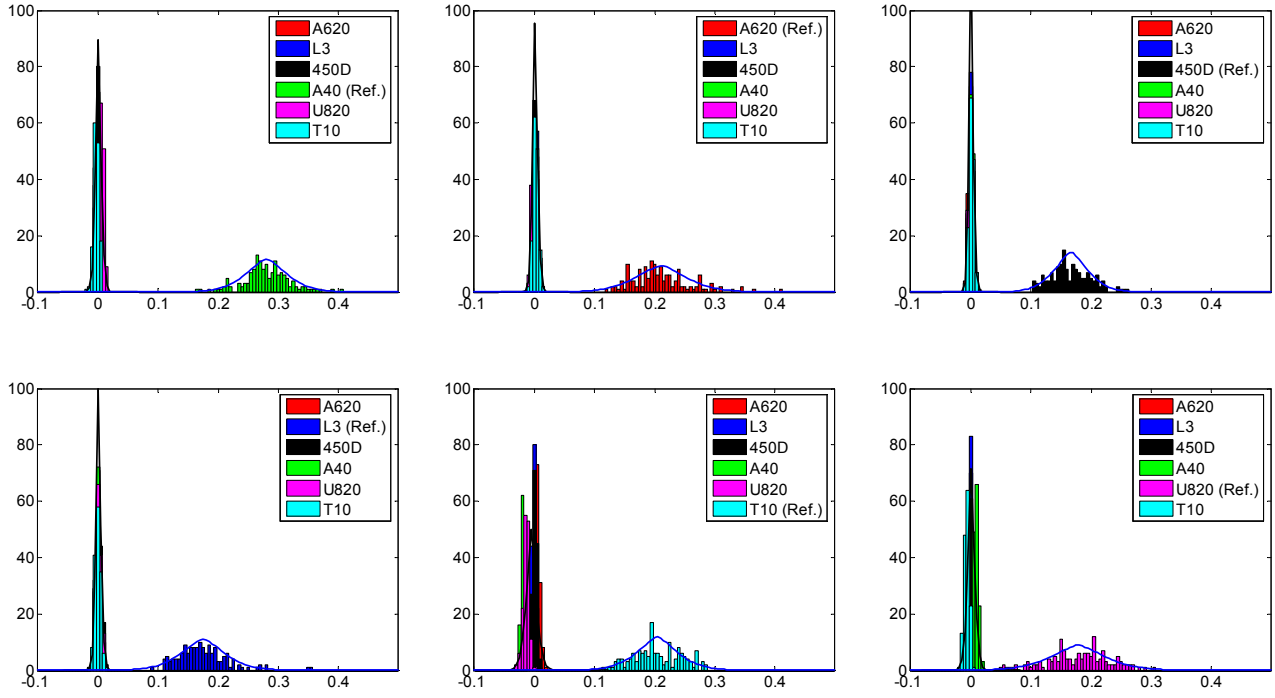

Figure 5. Results of Scheme 1.

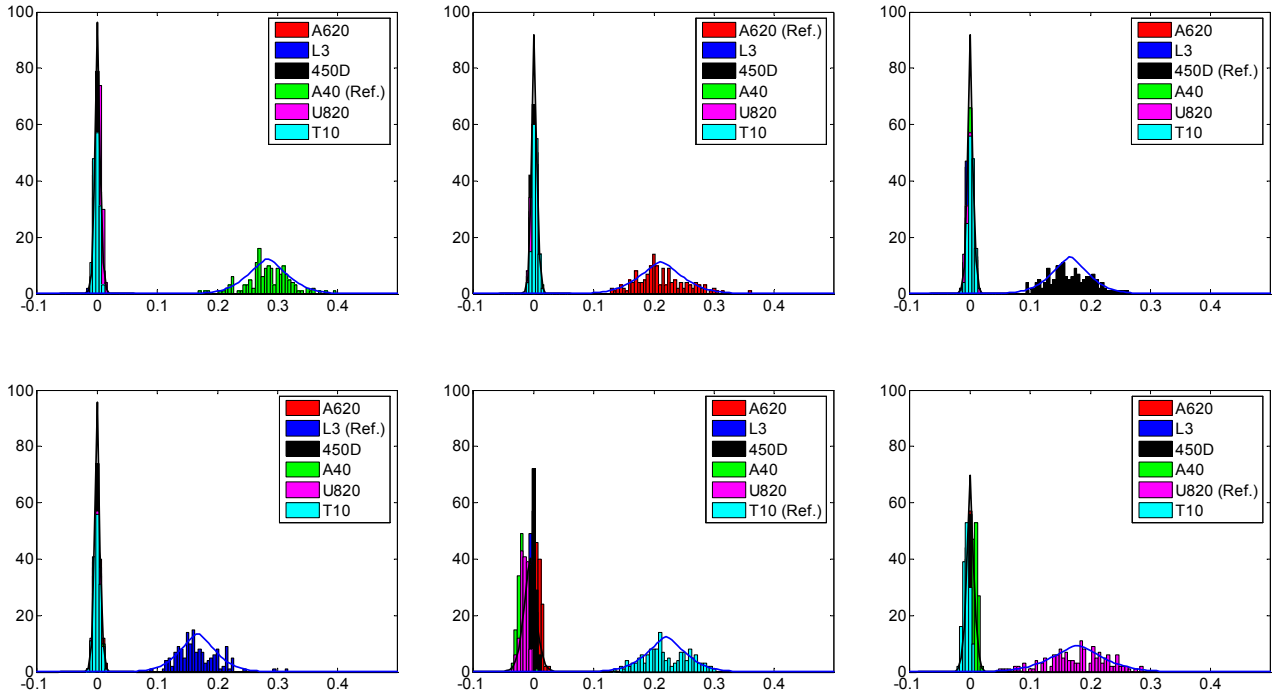

Figure 6. Results of Scheme 2.

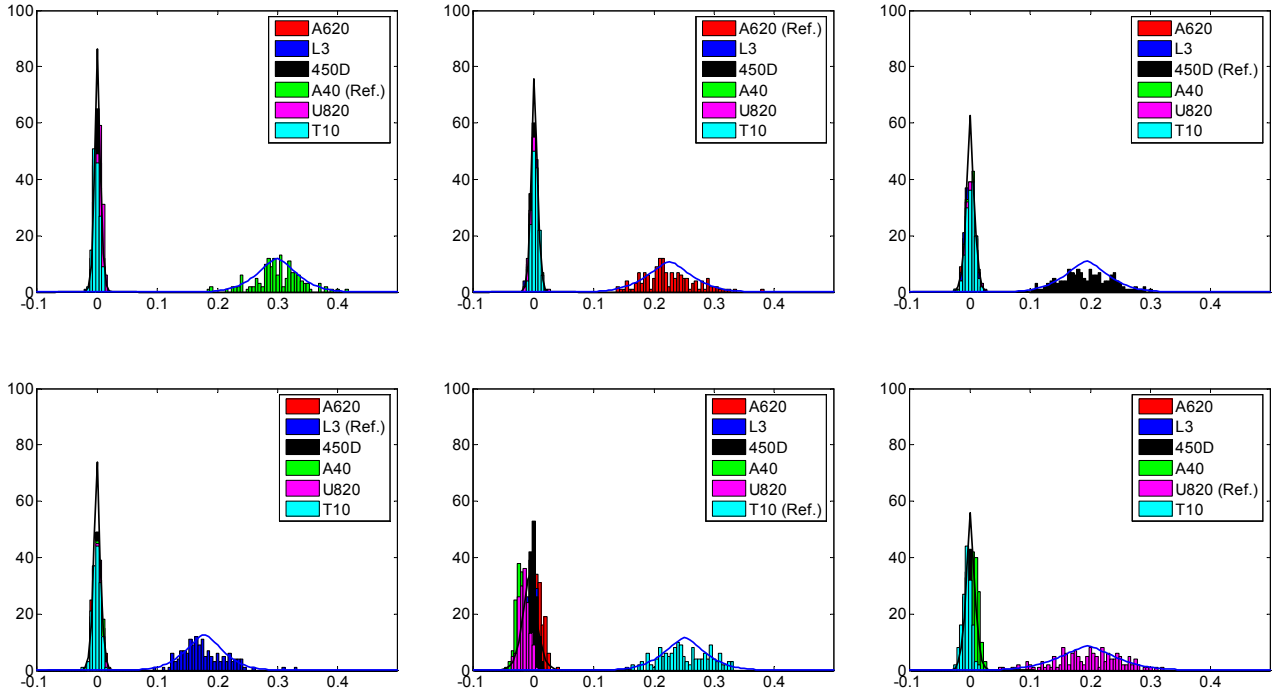

Figure 7. Results of Scheme 3.