Video Forgery Detection Using Correlation of Noise Residue

Chih-Chung Hsu [*1], Tzu-Yi Hung[*2], Chia-Wen Lin [*3], Chiou-Ting Hsu [#4]

[#] *Department of Electrical Engineering, National Tsing-Hua University*

101, Section 2, Kuang-Fu Road, Hsinchu, Taiwan 30013, R.O.C.

[1] d9661805@oz.nthu.edu.tw

[3] cwlin@ee.nthu.edu.tw

[#]*Department of Computer Science, National Tsing Hua University*

101, Section 2, Kuang-Fu Road, Hsinchu, Taiwan 30013, R.O.C.

[4] cthsu@cs.nthu.edu.tw

Abstract—**We propose a new approach for locating forged regions in a video using correlation of noise residue. In our method, block-level correlation values of noise residual are extracted as a feature for classification. We model the distribution of correlation of temporal noise residue in a forged video as a Gaussian mixture model (GMM). We propose a two-step scheme to estimate the model parameters. Consequently, a Bayesian classifier is used to find the optimal threshold value based on the estimated parameters. Two video inpainting schemes are used to simulate two different types of forgery processes for performance evaluation. Simulation results show that our method achieves promising accuracy in video forgery detection.**

I. INTRODUCTION

In recently years, due to the advances of network technologies, low-cost multimedia devices, sophisticated image/video editing software and wide adoptions of digital multimedia coding standards, digital multimedia applications have become increasingly popular in our daily life. However, the digital nature of the media files, they can now be easily manipulated, synthesized and tampered in numerous ways without leaving visible clues. As a result, the integrity of image/video content can no longer be taken for granted and a number of forensic-related issues arise.

Two types of forensics scheme are widely used for image/video forgery detection: active schemes and passive schemes. With the active schemes, the tampered region can be extracted using a pre-embedded watermark. However, this scheme must have source files to embed the watermark first; otherwise, the detection process will fail [1]. On contrary, the passive schemes extract some intrinsic fingerprint traces of image/video to detect the tampered regions.

When a real-world scene is captured by a digital camera, the information about the scene is processed by a pipeline of various camera components, such as color filter array (CFA), demosaicing, white-balancing, automatic gain control (AGC), Gamma correction, post-processing, and JPEG compression, before the final digital image is produced as shown in Fig. 1. In the imaging pipeline, each individual processing component modifies the input image via a particular processing algorithm, which may leave some intrinsic fingerprint traces out the output [1][2].

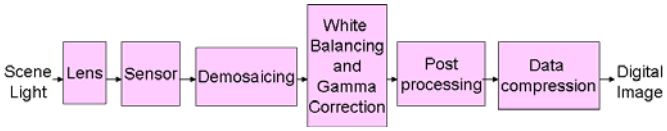


Fig. 1. Imaging pipeline.

In many digital image/video-based forensic applications, the intrinsic fingerprint traces, such as the process of CFA [3][4], Camera Response Function (CRF) [5], sensor pattern noise [6][7][8], and compression artefacts [9] can be used to detect image/video tampering such as resampling [10], copy and paste, slicing [4][11], and double compression [12]. Besides, the imperfect information in the camera is also used for forensics application [2].

Recently, sensor pattern noise has been successfully used as intrinsic sensor biometrics for nonintrusive forensic analysis [6][7][8]. For example, the sensor pattern noise has proven to be useful and reliable in identifying camera sources [6][8]. The method proposed in [6] first extracts the pattern noise images in training images captured with some specific cameras and the reference pattern noise image can then be obtained via averaging operation on these pattern noise images. The correlation measurement between the reference pattern noise image and pattern noise image is used here. The sensor pattern noise has also been used for scanner model identification and tampering detection of scanned images [8]. In [8], in addition to camera source identification, sensor pattern noise was first utilized for image forgery detection. This method proposed an accurate pattern noise extraction scheme. The above methods [6][7][8] need to pre-collect a number of images captured from specific video cameras to extract the sensor pattern noise of the cameras. Besides, it is difficult to extract sensor pattern noise from a video without a extensive variety of video contents.

This paper aims to address passive forgery detection in a digital video based on the statistical property of noise residue. We propose to analyze the temporal correlation of block-level noise residue to locate the tampered regions of a video. Our method does not need to pre-collect and pre-train the statistics

of noise residue for specific video cameras as the noise residue information can be easily on-the-fly extracted from the video to be authenticated. We propose to model the distributions of temporal noise correlation values of video blocks in forged and normal regions using a GMM model. In our method, the GMM model parameters are estimated using the Expectation-Maximization (EM) algorithm so that optimum thresholds can be derived accordingly using a Bayesian classifier. Two video inpainting schemes are used to evaluate the performance of the proposed method.

The rest of this paper is organized as follows. Section 2 presents the proposed video forgery detection scheme based on the noise residue of camera. In Section 3, the experimental results of the proposed schemes are demonstrated. Finally, Section 4 concludes this paper.

II. Proposed Video Forgery Detection Method

A. Overview of the proposed method

We propose a bottom-up approach for locating the forged/inpainted regions of a video based on block-level temporal noise correlation. Fig. 2 shows the flowchart of the proposed video forgery detection algorithm. In the first step, following the same process proposed in [4], the noise residue of each video frame is extracted by subtracting the original frame from its noise-free version. The wavelet denoising filter proposed in [17] is used to obtain the noise-free image.

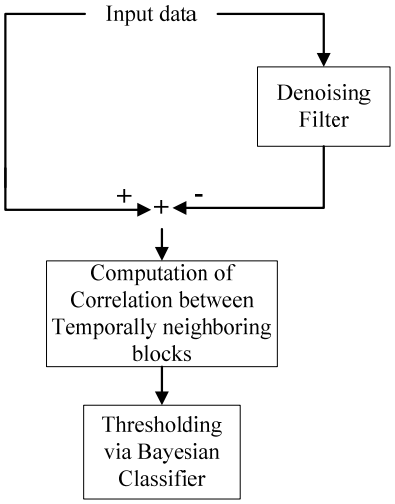


Fig. 2. Flowchart of our proposed method.

In the second step, each video frame is first partitioned into non-overlapping blocks of size $N \times N$. The correlation of the noise residue between the same spatially indexed blocks of two consecutive frames is then computed as illustrated in Fig 3.

The final step is to locate tampered blocks by analyzing the statistical properties of block-level noise correlations. In the first part of this step, a simple thresholding scheme is exploited to obtain a coarse classification. Based on the coarse classification, a GMM model is applied to characterize the statistical distributions of block-level temporal noise correlations for tampered and non-tempered regions, respectively. The GMM model parameters are then estimated using the EM algorithm so that optimum thresholds can be derived accordingly using the maximum-likelihood (ML) estimation and Bayesian classifier. The detail of each step is elaborated in the following.

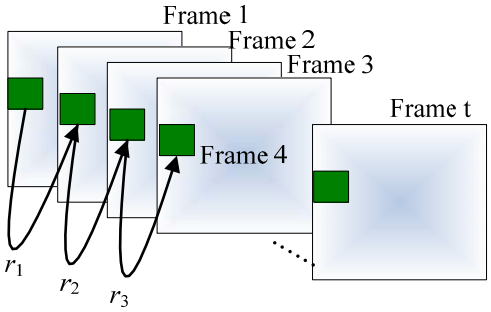


Fig. 3. Illustration of computing the correlations of the noise residue between every two temporally neighboring video blocks.

B. Extraction of Noise Residue

As mentioned above, we adopt the same denoising filter proposed in [4][17]. First, we assume that the high-frequency wavelet coefficients can be modeled as the sum of a stationary white Gaussian noise and a noise-free image. The denoising process is composed of four steps as summarized below:

1. A four-level wavelet decomposition is performed on a noisy image to obtain its wavelet coefficients. After the decomposition, only high-frequency components (i.e., the LH $h(i,j)$, HL $v(i,j)$ and HH $d(i,j)$ subbands) are used for processing.
2. Second, the local variance of each wavelet coefficient is estimated. For each wavelet coefficient, we define a window size of $W \times W$, where $W \in \{3,5,7,9\}$. The local variance is computed by

$$\hat{\sigma}_W^2(i,j) = \max\left(0, \frac{1}{W^2}\sum_{(i,j)\in N} c^2(i,j) - \sigma_0^2\right) \quad (1)$$

   where $\sigma_0 = 5$ $c(i,j)$ denotes the wavelet coefficient in each sub-band ($h(i,j)$, $v(i,j)$, and $d(i,j)$). Then, we take the minimum value among the local variances by

$$\sigma^2(i,j) = \min\left(\hat{\sigma}_3^2(i,j), \hat{\sigma}_5^2(i,j), \hat{\sigma}_7^2(i,j), \hat{\sigma}_9^2(i,j)\right) \quad (2)$$

3. The wiener filter with the following profile is used for denoising.

$$c_{\text{den}}(i,j) = c(i,j)\frac{\hat{\sigma}^2(i,j)}{\hat{\sigma}^2(i,j) + \sigma_0^2} \quad (3)$$

4. For each wavelet coefficient, repeat previous steps until the process converged. Finally, the inverse wavelet transform is used to obtain the noise-free image.

After the noise-free image is obtained, the noise residual $n(i,j)$ can be easily extracted by subtracting the original image from its noise-free version. The noise residue $n(i,j)$ consists PNU and high-spatial-frequency details of image content. In [4], the PNU was extracted from noise residual $n(i,j)$ via an averaging operation. A more sophisticated and accurate sensor pattern noise extraction scheme was proposed in [8]. It however, consumes significantly more computation. Although sensor pattern noise can be used as an intrinsic fingerprint to effectively identify image/video forgery. It is in general difficult to extract sensor pattern noise from a video without an extensive variety of video contents such as in a surveillance video. In this work, we use the extracted noise residue $n(i,j)$ directly to find the forged regions. Although, in addition to sensor pattern noise, the noise residue also contains high-frequency image details, it is still useful for video forgery detection.

C. Calculation of Block-level Noise Correlation Values

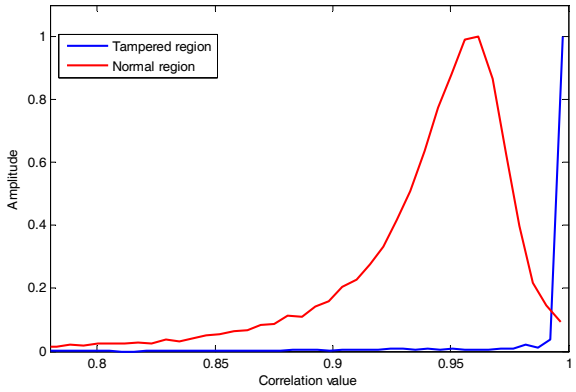
Let $n_{i,j}$ denote the noise residual at pixel coordinate (i,j). The correlation value r between previous frame and current frame on each block (shown in Fig. 4) can be defined as

$$r = \frac{\sum_i \sum_j \left(n_{i,j}^t - \bar{n}^t\right)\left(n_{i,j}^{t-1} - \bar{n}^{t-1}\right)}{\sqrt{\sum_i \sum_j \left(n_{i,j}^t - \bar{n}^t\right)^2 \sum_i \sum_j \left(n_{i,j}^{t-1} - \bar{n}^{t-1}\right)^2}} \quad (4)$$

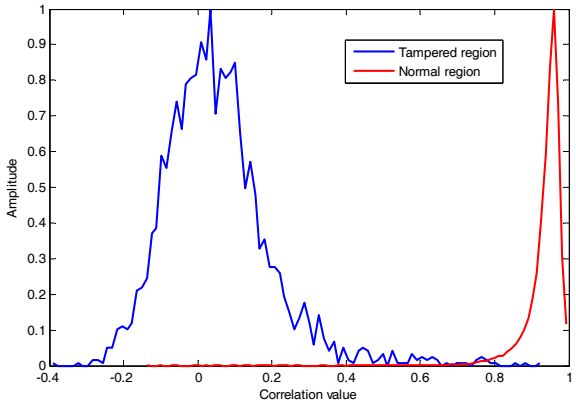
where t denotes the t-th frame and $\bar{t}^n$ is the mean value of the noise residual at t-th frame.

When a region is forged, the correlation value of temporal noise residue in the region is usually changed (increased or decreased) depending on the forgery scheme used. Fig 4 shows the histograms of block-level correlation values of every two consecutive frames for a forged still-background surveillance video. The red curves indicate the distributions of non-forged blocks, whereas the blue ones indicate those of forged blocks. Two different video inpainting schemes are used to simulate two most representative kinds of forgery schemes that lead to contrary effects on noise correlation. The first is temporal copy-paste inpainting which find the most coherent block from to fill in the tampered region. Obviously the correlations in the forged region become higher, almost approaching unity, since in a video with still background the content selected for forging temporally neighboring blocks are usually the same to maintain the temporal coherence of the forged region.

The other is the example-based texture synthesis scheme proposed in [14] that fills in a region from sample textures. It is one of the state-of-the-art image inpainting schemes. Although texture synthesis is in general not suitable for inpainting a video since it is difficult to maintain temporal coherence between successive frames after inpainting, we use it to simulate the tampering processes that will in effect decorrelate the sensor pattern noise as the sample textures for completing a region are selected from different locations, thereby reducing the temporal correlation of noise residue. Note, using the temporal copy-paste inpainting to complete a removed object with a moving background or dynamic scene (e.g., a video captured with a moving camera) may also lead to similar effect since it fill in a region using content with different noise patterns compared to those of the temporally neighboring regions. The two inpainting schemes are used in this work to simulate two typical kinds of tampering processes for evaluating the performance of the proposed forgery detection schemes.



(a)



(b)

Fig. 4. Comparison of distributions of noise correlation values between two temporally neighboring video blocks in forged and non-forged regions. The forging schemes used are: (a) temporal copy-paste inpainting [13], and (b) example-based texture synthesis [14].

D. Forgery Detection by Statistical Analysis of Noise Residue

Since the tampering process usually changes the temporal statistical property of sensor residue, we can distinguish the tampered regions from the non-tampered ones by analyzing the statistical properties of block-level noise correlation.

After the correlation values of every two temporally neighboring blocks are obtained, the parameters of the distributions of normal blocks and forged blocks are estimated respectively via the maximum-likelihood estimation. For a Gaussian distributed signal, the parameters can be estimated easily by calculating

$$\bar{r} = \sum_{n=1}^{N} r_n \quad (5)$$

and

$$\sigma_r^2 = \frac{1}{N}\sum_{n=1}^{N}(r_n - \bar{r})^2 \tag{6}$$

In typical applications, in a forged video, the area of forged regions is usually much smaller than that of normal region. Based on this assumption, we propose a simple pre-classification scheme to quickly determine whether a video frame has been forged without resorting to computation-intensive fine classification for non-tampered video frames, thereby achieving significant computation saving. Besides, the pre-classification result can be used to speed up the model adaptation in the fine-classification. The pre-classification is defined as follows:

$$Class_n = \begin{cases} 0 & |r_n - \bar{r}| < k \cdot \sigma_r \\ 1 & \text{otherwise} \end{cases} \tag{7}$$

where $Class_n$ denotes the binary classification mask for the n-th block with a value of 1 indicating the block has been forged.

In the pre-classification stage, when the percentage of tampered blocks detected in a video frame exceeds a predetermined threshold, a fine-classification process is performed to refine the detection result. Otherwise, the video frame is classified as a non-tampered frame, and no further detection process is preformed. Based on the pre-classification, a GMM model is applied to characterize the statistical distributions of block-level noise correlations for the tampered and normal regions, respectively. The GMM model parameters are then estimated using the EM algorithm [19] so that optimum thresholds can be derived accordingly using the maximum-likelihood (ML) estimation and Bayesian classifier. The means and variances of the block-level noise correlation of the forged and normal classes are used as the initial values in the EM algorithm to speed up the iteration process.

For simplicity but without loss of generality, we assume that there are two Gaussian distributions in a forged video. For the two-class problem, the discriminate function can be defined as

$$g_i(r) = -\frac{1}{2}\ln 2\pi + \ln \sigma_r - \frac{(r-\bar{r})^2}{2\sigma_r^2} + \ln P(\omega_i) \tag{8}$$

where $P(\omega_i)$ denotes the prior of the ith class, which can be approximated by the result of pre-classification. To obtain the classification hyper-plane, we compute the threshold value of r by setting equation $g_1(r) - g_2(r) = 0$ which is the optimal threshold value with minimum classification error according to the Bayesian classification theory.

III. Experimental Results

In our experiments, three 200-frame surveillance-like test sequences with still backgrounds were captured using a JVC GZ-MG50TW digital camcorder with a frame rate of 30 fps. The built-in video encoder of the video camera is MPEG-2. The resolution of each frame is ITU-R 601 (720×480) and the bit-rate is 8.5 Mbps. Therefore the compression ratio is about 15:1 for the 4:2:0 Y-Cb-Cr color format, and 30:1 for the full color format.

As mentioned above, the temporal copy-paste inpainting [13] and example-based texture synthesis [14] were used for forging the human objects in the videos. As shown in Table I, the forgery detection performance is evaluated by the precision and recall rates, which are calculated from the ground-truths as defined below:

$$\text{Precision} = \frac{N_{\text{hit}}}{N_{\text{hit}} + N_{\text{false}}} \tag{9}$$

$$\text{Recall} = \frac{N_{\text{hit}}}{N_{\text{hit}} + N_{\text{miss}}} \tag{10}$$

where N_{hit} represents the number of correct detections, N_{false} denotes the number of false positives, and N_{miss} denotes the number of misses.

TABLE I
Performance Evaluation of the Proposed Forgery Detection Scheme for Three Test Sequences

	Recall	Precision	Miss_rate	False_positive_rate
Temporal Copy-Paste Inpainting				
Seq_1	62.54%	98.22%	37.46%	1.78%
Seq_2	40.76%	93.28%	59.24%	6.72%
Seq_3	64.00%	98.34%	36.00%	1.66%
Example-Based Texture Synthesis Inpainting				
Seq_1	76.31%	94.83%	23.69%	5.17%
Seq_2	72.86%	90.78%	27.14%	9.22%

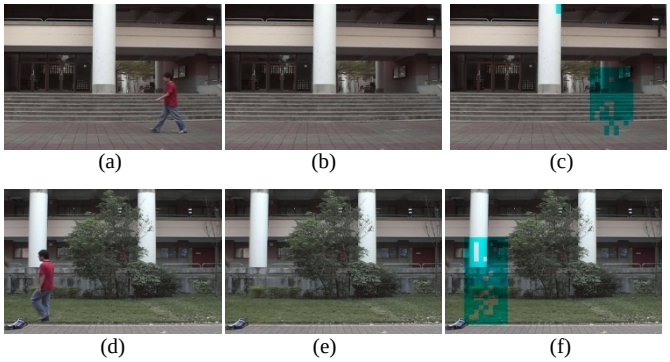


Fig. 6. Snapshots of two test sequences for the temporal copy-paste inpainting scheme: (a),(d) original frames, (b),(e) the inpainted frames, and (c),(f) the corresponding detection result.

The experimental results show that, for both two video inpainting attacks, the proposed method achieves high precision rates with rather small false positive rates, where the false positives are mainly due to too high or too low illumination which leads to low noise residual energy. Besides, the noise residual is also content dependent making the correlation feature not very stable for applications with dynamic scene such as a video captured with a moving camera. The detection result for texture-synthesis-based inpainting achieves better recall rates because the noise correlation value drops drastically after inpainting. Fig. 6 shows some snapshots of the original video frames, their inpainted frames using the two inpainting schemes, and the corresponding detection results using the proposed method. The green blocks indicate

the forged regions detected. The test videos, their inpainted versions, and the detection results can be found in [20].

Although the noise correlation is fairly reliable feature in fine-quality video, it is sensitive to quantization noise. We used an encoder to recompress the test sequences with different settings of quantization step-sizes (with compression ratios ranging from 27 to 62) to evaluate the effect of compression on detection accuracy. Fig. 7 shows that the higher the compression ratio, the lower the detection precision, as much of noise may disappear after coarsely quantizing DCT coefficients. Therefore the proposed noise correlation is not reliable for low-quality video such as low-bandwidth Internet streaming videos. However, quantization noise itself can be used as an intrinsic fingerprint signature for forgery detection in low-bit-rate compressed video since quantization noise is a compression-oriented feature that appears in the form of blocking effects. For example, a method of using blocking effects to detect image forgery was proposed in [9]. Adequately combining the noise residue feature with compression-oriented features such as blocking effects may improve the reliability and accuracy of forgery detection in a wide range of video bit-rates.

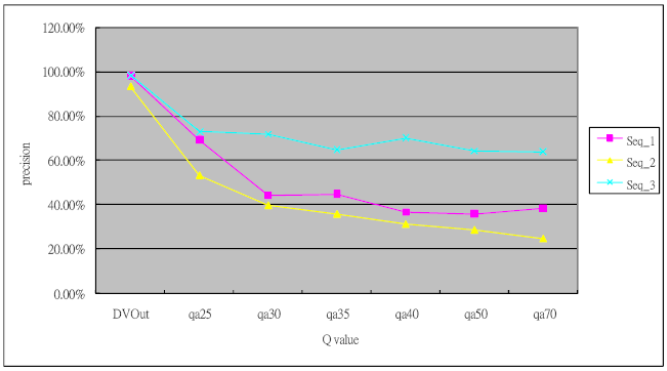


Fig. 7. Comparison of detection precision for with different quantization step-size settings (with compression ratios ranging from 27 to 62).

IV. Conclusion

In this paper, we have proposed a digital video forgery detection scheme using temporal noise correlation without the need of embedding any prior digital signature in the compressed video. We have also proposed a statistical classification scheme based on a GMM model and the Bayesian classifier. In our experiments, two video inpainting schemes are used to simulate two different types of tampering processes for performance evaluation. Experimental results show that the proposed method achieves promising detection accuracy for fine-quality videos.

Acknowledgment

This work was supported in part by the Ministry of Economic Affairs (MOEA), Taiwan, R.O.C., under grant 96-EC-17-A-02-S1-032.

References

[1] T.-T. Ng, S.-F. Chang, C.-Y. Lin, and Q. Sun, "Passive-blind image forensics", *In Multimedia Security Technologies for Digital Rights*, W. Zeng, H. Yu, and C.-Y. Lin (eds.), Elsevier, 2006.

[2] A. Swaminathan, M. Wu, and K. J. R. Liu, "Digital image forensics via intrinsic fingerprints," *IEEE Trans. Information Forensics and Security*, vol.3, no.1, pp.101-117, Mar. 2008.

[3] A.C. Popescu and H. Farid, "Exposing digital forgeries in color filter array interpolated images," *IEEE Trans. Signal Process.*, vol. 53, no.10, pp. 3948-3959, Oct. 2005.

[4] S. Bayram, H. T. Sencar, and N. Memon, "Source camera identification based on CFA interpolation," in *Proc. IEEE Int. Conf. Image Processing*, vol.3, no., pp. III-69-72, 11-14, Sept. 2006.

[5] Y.-F. Hsu and S.-F. Chang, "Image splicing detection using camera response function consistency and automatic segmentation," *in Proc. IEEE Conf. Multimedia Expo.*, pp. 28-31, July 2007, Beijing, China.

[6] J. Lukáš, J. Fridrich, and M. Goljan, "Digital camera identification from sensor pattern noise," *IEEE Trans. Information Forensics Security*, vol.1, no.2, pp. 205-214, June 2006.

[7] J. Lukáš, J. Fridrich, and M. Goljan, "Detecting digital image forgeries using sensor pattern noise," in *Proc. SPIE Electronic Imaging, Photonics West*, pp. 60720Y-1 – 11, Jan. 2006.

[8] M. Chen, J. Fridrich, and J. Lukáš, "Determining image origin and integrity using sensor pattern noise," *IEEE Trans. Information Forensics Security*, vol.3, no.1, pp. 74-90, Mar. 2008.

[9] S. Ye, Q. Sun and E.-C. Chang, Detecting digital image forgeries by measuring inconsistence of blocking artifact, July 2007, Beijing, China.

[10] A.C. Popescu and H. Farid, "Exposing digital forgeries by detecting traces of re-sampling," *IEEE Trans. Signal Process.*, vol. 53, no.2, pp. 758-767, Feb. 2005.

[11] Y. Q. Shi, C. Chen, and W. Chen, "A natural image model approach to splicing detection," in *Proc. ACM Multimedia Security Workshop*, pp. 51-62, Sept. 2007, Dallas, Texas.

[12] Z. Fan and R. L. de Queiroz, "Identification of bitmap compression history: JPEG detection and quantizer estimation," *IEEE Trans. Image Process.*, vol. 12, no. 2, pp. 230–235, Feb. 2003.

[13] K. A. Patwardhan, G. Sapiro, and M. Bertalmio, "Video inpainting under constrained camera motion," *IEEE Trans. Image Process.*, vol.16, no. 2, pp. 545-553, Feb. 2007

[14] A. Criminisi, P Perez, and K. Toyama, "Region filling and object removal by exemplar-based image inpainting," *IEEE Trans. Image Process.*, vol.13, no.9, pp. 1200-1212, Sept. 2004.

[15] G. C. Holst, *CCD Arrays, Cameras, and Displays*, 2nd ed. Winter Park, FL, and Bellingham, WA: JCD & SPIE, 1998.

[16] J. R. Janesick, *Scientific Charge-Coupled Devices*, SPIE vol. PM83, Bellingham, WA, 2001.

[17] M. K. Mihcak, I. Kozintsev, and K. Ramchandran, "Spatially adaptive statistical modeling of wavelet image coefficients and its application to denoising," *in Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, vol. 6, pp. 3253–3256, Mar. 1999, Phoenix, AZ.

[18] *Application Note, CCD Image Sensor Noise Sources, Kodak: Image Sensor Solutions*, Jan. 2005.

[19] T. K. Moon and W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing*, Prentice Hall, Upper Saddle River, NJ, 1999.

[20] NTHU Forensics project. [Online]. Available: http://www.ee.nthu.edu.tw/cwlin/forensics/forensics.htm