TAMPERING IDENTIFICATION USING EMPIRICAL FREQUENCY RESPONSE

Wei-Hong Chuang, Ashwin Swaminathan, and Min Wu
Department of Electrical and Computer Engineering, University of Maryland, College Park

ABSTRACT

With the widespread popularity of digital images and the presence of easy-to-use image editing software, content integrity can no longer be taken for granted, and there is a strong need for techniques that not only detect the presence of tampering but also identify its type. This paper focusses on tampering-type identification and introduces a new approach based on the Empirical Frequency Response (EFR) to address this problem. We show that several types of tampering operations, both linear shift invariant (LSI) and non-LSI, can be characterized consistently and distinctly by their EFRs. We then extend the approach to estimate the EFR for scenarios where only the final image is available. Theoretical reasoning supported by experimental results verify the effectiveness of this method for identifying the type of a tampering operation.

Index Terms— Multimedia forensics, tampering type identification.

1. INTRODUCTION

Nowadays, due to the widespread popularity of digital cameras and online photo hosting services, a large number of photographs have been generated and distributed. At the same time, the advent of various image editing software packages has made altering the photo content easier even for novice users. Since the authenticity of digital photos impacts on how we use it, content integrity has become an important forensic issue. For a given photo, one may ask if it has been tampered or manipulated and further by what *type* of tampering operation. This paper focuses on the latter question and presents a framework to determine the *type* of tampering operation that has been performed.

Prior works fall into two main categories. In the first category, methods have been proposed to detect resampling [1], JPEG compression [2], and Gamma correction [3], by extracting certain salient features that would help distinguish such tampering from unprocessed images. Although these methods can be employed to identify the type and the parameters of the tampering operation, an exhaustive search over a pool of operations is required to detect tampering and to identify the type of tampering operation. Therefore, there is a strong need for universal technique to detect and identify tampering.

In the second category, classifier-based approaches to detect image tampering were proposed in [4][5], where features based on analysis of variance [4] and higher order wavelet statistics [5] have been used. In [6], a framework was proposed by modeling tampering as a combination of a linear and shift-invariant (LSI) and a non-LSI part. The authors present methods to estimate the LSI part of manipulation operation and compare the estimate to an identity transform to detect tampering. These works aim to just *detect* tampering and thus focus on answering whether the given image was tampered or not, and are not for identifying the *type* of tampering.

In this work, we propose a framework based on the Empirical Frequency Response (EFR) that aims to identify the manipulation type. We show that many classes of LSI or non-LSI image processing operations, such as resampling, JPEG compression, and non-linear filtering, exhibit distinctive patterns in their EFRs. Theoretical reasoning supported by experimental results also verifies the effectiveness of this method for identifying the type of a tampering operation.

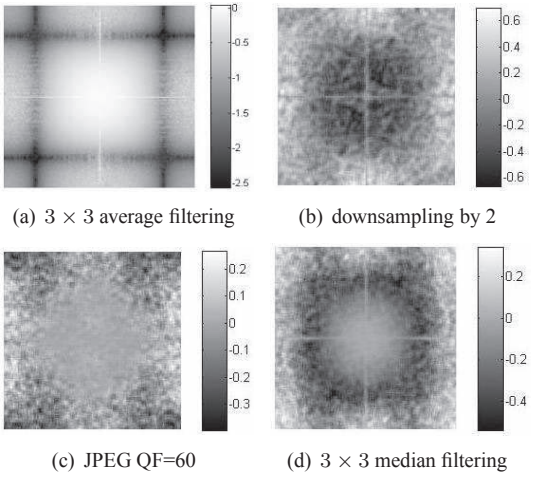
This paper is organized as follows. We define the Empirical Frequency Response (EFR) in Section 2 and show distinctive EFRs. The results on using the EFR as a tampering analysis tool are discussed in Section 3. Since the EFR is, in fact, not readily available in practice, we discuss methods to estimate EFR in Section 4 just based on the output image, and propose approaches to improve the accuracy. We conclude this paper in Section 5.

2. EMPIRICAL FREQUENCY RESPONSE FOR IDENTIFYING TYPE OF TAMPERING

It is well known that linear and shift-invariant (LSI) systems can be characterized by their frequency responses. For example, a 3×3 average filter has a 2-D sinc-like frequency response as shown in Fig. 1(a) and the frequency response of an identity system whose output equals to the input is flat. However, image processing operations are often non-LSI and input-independent frequency response is not defined for such systems. In this paper, we represent such manipulations using the Empirical Frequency Response (EFR) [7]. For different types of tampering, we show that the EFR is consistent and can therefore be employed to identify manipulation type.

The EFR of a system $H_X(\boldsymbol{\omega})$ is defined as the ratio of the Fourier transform of the system output $Y(\boldsymbol{\omega})$ and the Fourier transform of the input $X(\boldsymbol{\omega})$, *i.e.*, $H_X(\boldsymbol{\omega}) = \frac{Y(\boldsymbol{\omega})}{X(\boldsymbol{\omega})}$. In case of digital images, we replace the Fourier transform by discrete Fourier Transform (DFT), but the idea of EFR remains the same. The EFR is input-dependent for non-LSI systems, and when the system is LSI, it coincides with the frequency

Email contact: {whchuang, ashwins, minwu}@umd.edu



(a) 3×3 average filtering (b) downsampling by 2

(c) JPEG QF=60 (d) 3×3 median filtering

Fig. 1. Typical EFRs for four different manipulations. The EFR is shown in a log scale with the center part representing the low-frequency region.

response. Fig. 1 illustrates typical EFRs for different manipulations including (i) downsampling by 2 (denoted by $\downarrow 2$; the notation $\uparrow$ is similarly for upsampling); (ii) JPEG compression with quality factor (QF) 60, and (iii) 3×3 median filtering (a popular non-linear filter). We obtain similar or "consistent" EFRs for a majority of photographs in our database; this suggests that even though the EFRs are signal dependent for non-LSI systems, the differences are often minor and similar manipulations produce similar EFRs. In the following, we analyze the reasons behind this consistency for operations such as resampling, JPEG compression, and median filtering.

2.1. EFR Consistency for Resampling Operations

Manipulations such as resampling change the original image structure, especially among camera-captured images. This is because most digital cameras adopt a color filter array (CFA) to capture the information about the real-world scene. The CFA consists an array of color sensors, each of which captures a corresponding color of the real-world scene at an appropriate pixel location. After sampling, only one color is recorded at each pixel location, and interpolation is performed to obtain the remaining color components.

This implicit structure from sampling and interpolation would be destroyed by resampling. For illustration, let us consider a 1-D case with $x(n)$ denoting the interpolated camera output, and $y(n)$ representing the result after downsampling $x(n)$ by 2. We assume that $x(n)$ is interpolated such that each odd entry in $x(n)$ is the linear combination of its two neighboring entries, *i.e.*, $x(2n+1) = 1/2[x(2n) + x(2n+2)]$. Under this assumption, we can derive the EFR to be

$$H_X(\omega) = \underbrace{\frac{1}{1+\cos\omega}}_{\textit{camera factor}} \underbrace{\frac{B(\omega)}{B(2\omega)}}_{\textit{content factor}} . \qquad (1)$$

Here, $B(\omega)$ denotes the discrete time Fourier transform (DTFT) of $b(n) = x(2n)$. Since $1+\cos\omega \geq 1$ for $0 \leq \omega \leq \frac{\pi}{2}$ and $0 \leq 1+\cos\omega \leq 1$ for $\frac{\pi}{2} \leq \omega \leq \pi$, the magnitude of the "camera factor" in (1) is smaller for $0 \leq \omega \leq \frac{\pi}{2}$ and is larger for ω close to π. In contrast to the camera factor in (1), the content factor in (1) is dependent on the input signal. For typical values of $b(n)$ sampled from a natural images, the content factor is bounded and follows a similar trend across different images. Therefore, $H_X(\omega)$ is primarily determined by the signal-independent camera factor and, thus is consistent across a gamut of natural images. This analysis also reveals that for most CFA-interpolated photographs, higher and lower bands in the EFR would be strengthened and weakened, respectively, after direct downsampling. Such changes can be observed in Fig. 1(b).

Resampling by a general L/M factor can also be analyzed in a similar manner. In this case, we can decompose the resampling operation into the cascade of an upsampler, $\uparrow L$, a low-pass filter $F(\omega)$, and a downsampler, $\downarrow M$, and the EFR can be derived to be

$$H_X(\omega) \approx \underbrace{\frac{1}{M} F\left(\frac{\omega}{M}\right) \frac{1+\cos(\frac{L\omega}{M})}{1+\cos\omega}}_{\textit{camera factor}} \underbrace{\frac{B(\frac{2L\omega}{M})}{B(2\omega)}}_{\textit{content factor}}, 0 \leq \omega \leq \pi.$$

The EFR is this case can be again decomposed into signal-independent and signal-dependent factors. By an argument similar to that for direct downsampling by 2, we can show that EFR of a general resampling operation is consistent.

2.2. EFR Consistency for JPEG and Median Filtering

JPEG compression is done through quantization of the block-based (usually 8×8 or 16×16) discrete cosine transform (DCT) coefficients. Because the quantization steps for low-frequency coefficients are usually small, JPEG compression tends to preserve the low-frequency components, but for high-frequency bands, large quantization steps have the effect of destroying image datails. The blocking effect causes discontinuity across block boundaries, particularly in the vertical or horizontal directions. Combining these factors, the EFR of JPEG compression is expected to have values close to 1 (or 0 in the log scale) in the low-low frequency region, smaller values in high-high frequency bands due to loss of image detail, and larger values in the low-high and high-low bands due to blocking artifacts as is observed experimentally in Fig. 1(c).

Median filtering is known to have a frequency response similar to that of an average filter for frequencies lower than $2\pi/\alpha$, where α is the filter order [7]. Outside this region, some high-frequency components are retained to preserve the signal sharpness and some others are weakened as shown in the example in Fig. 1(d). In the next section, we build upon our observation on the EFR consistency across different tampering operations and present a framework for determining the type of tampering operations.

3. TAMPERING OPERATION ANALYSIS USING EFR

Experiment Setup and Feature Selection: In this section, we study the performance of EFR in characterizing different types of tampering operations. As demonstrated in Section 2, the EFR is a function of the tampering operation, the camera used, and to some extent dependent on the nature of the input image for non-LSI systems. For example, in the case of resampling, color interpolation coefficients and the low-pass filter are usually a function of the camera and may vary among different camera models. In order to take the effect of the camera into consideration, we employ a data set containing six cameras (Canon PowerShot A75, FujiFilm FinePix S3000, Sony CyberShot DSC P72, Minolta DiMage S304, Epson PhotoPC 650, and Fujifilm FinePix F31fd) in our experiments with forty photographs from each camera.

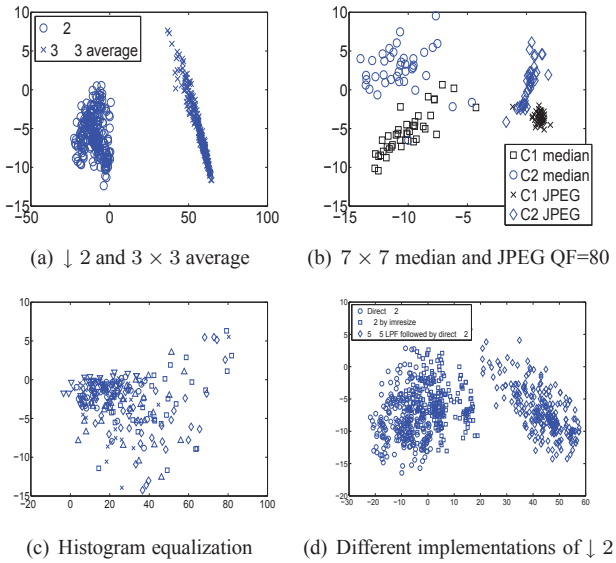
We consider 16 types of manipulations, including resampling, LSI filtering, non-LSI filtering, compression, and a representative point operation. The settings are: (O1) $\downarrow 2$, (O2) $\downarrow 4$, (O3) $\downarrow 2$ by the MATLAB function imresize with default parameters, (O4) $\uparrow 2$ by imresize, (O5) $\uparrow 1.5$ by imresize, (O6-O7) 3×3 and 5×5 average filtering followed by $\downarrow 2$, (O8-O11) 3×3, 5×5, 7×7, and 9×9 average filtering, (O12-O13) 3×3 and 7×7 median filtering, (O14-O15) JPEG QF=60 and 80, (O16) histogram equalization.

We compute EFRs by extracting two corresponding 256×256 blocks from the input and output images, and use the fixed-sized discrete Fourier transform (DFT) to approximate the DTFT. For resampling operations that change the image sizes, we apply appropriate zero padding in spatial domain to interpolate the frequency components. We pre-process the EFRs by average filtering to reduce the effects of noise, and reduce their dimensionality by first donwsampling it to size 64×64 and then by applying Principal Component Analysis (PCA) to produce 8 features per image. The dimension 8 is selected experimentally; the performance begins to degrade as more features are included.

Consistency among EFRs: Fig. 2 plots the 2-D principal-component projections of the EFRs for different tampering operations. We notice from Fig. 2(a) that operations such as $\downarrow 2$ and 3×3 average filtering exhibit strong inner-operation consistency with the features forming very tight clusters.

The effect of cameras can be studied by capturing the same content using different cameras and examining the consistency of the EFR for different tampering operations. Fig. 2(b) shows the 2-D projections of EFRs from two cameras of two post-camera manipulations, namely, 7×7 median filtering and JPEG compression with quality factor 80. We see from the figure that the features form four small clusters, but those which belong to the same operation are much closer. This is another level of EFR consistency but still justifies our choice of employing EFRs for tampering type identification.

The EFRs of some operations may have higher dependences on the inputs and thus can be expected to have larger



(a) $\downarrow 2$ and 3×3 average (b) 7×7 median and JPEG QF=80

(c) Histogram equalization (d) Different implementations of $\downarrow 2$

Fig. 2. Plots showing the 2-D projection of the EFR for (a) downsampling and average filtering, (b) median filtering and JPEG compression across two different cameras (denoted by Cam1 and Cam2), (c) histogram equalization for which different markers represent different cameras, and (d) different implementations of $\downarrow 2$.

variations; histogram equalization shown in Fig. 2(c) is a representative example. Such variation reduces the consistency in EFR. Nevertheless, the EFR can be employed for identifying the manipulation type as long as the EFRs of considered operations do not occupy the same region in the feature space.

We also find that the EFRs in the feature space can reveal the intrinsic operation similarities among operations. For example, Fig. 2(d) compares three possible implementations of $\downarrow 2$: (O1), (O3) and (O7), which shows that (O3) is closer to (O1) rather than (O7), suggesting that imresize is more similar to direct $\downarrow 2$ and the low-pass filtering is not obvious.

Identifying Tampering Type using EFR: Now, we examine the classification performance using the EFR as features. We employ Gaussian Mixture Model (GMM) to learn each category and classify the EFR based features using a Maximum-Likelihood (ML) approach. We randomly employ thirty training photographs from each camera for training the classifier and use the remaining photographs for testing, repeating this process for twenty times to obtain the average performance.

We group the 16 operations into 6 categories: (C1) $\downarrow 2$, $\downarrow 2$ by imresize, $\downarrow 4$, 3×3 and 5×5 average followed by $\downarrow 2$, (C2) $\uparrow 2$ and $\uparrow 1.5$ by imresize, (C3) 3×3, 5×5, 7×7, 9×9 average, (C4) 3×3 and 7×7 median filtering, (C5) JPEG QF=60 and 80, and (C6) histogram equalization. Such a grouping partitions operations into categories consistent with signal-processing knowledge. Table 1 shows the classification performance using the 2-component ML-GMM approach. The average accuracy is 95.9% suggesting that the EFR can discriminate between types of tampering operations.

Table 1. Confusion matrix with the Original EFR.

%	1	2	3	4	5	6
1	93.2	0.6	0.3	5.2	0.5	0.3
2	3.1	96.1	0.0	0.1	0.7	0.0
3	1.2	0.0	98.8	0.0	0.0	0.0
4	6.7	0.0	0.0	92.2	0.6	0.4
5	1.9	0.3	0.0	1.3	95.5	1.0
6	0.2	0.0	0.0	0.1	0.1	99.7

Table 2. Confusion matrix with the Estimated EFR.

%	1	2	3	4	5	6
1	79.4	1.1	0.4	6.9	4.0	8.2
2	3.5	95.7	0.0	0.7	0.2	0.0
3	0.6	2.8	96.3	0.2	0.0	0.0
4	5.6	0.0	0.0	89.5	0.5	4.4
5	7.6	0.0	0.1	1.6	75.4	15.3
6	6.0	0.0	0.0	10.3	12.1	71.7

4. ESTIMATING EFR BY BLIND DECONVOLUTION

In most applications involving tampering detection, we do not have access to the camera output (namely, the system input) and the EFR of the system cannot be readily determined. To address this problem, in our work, we estimate the LSI component of the EFR using the iterative blind deconvolution procedure described in [6], which is only briefed here due to space limit. The iterative blind deconvolution approach works by repeatedly applying known constraints in the pixel domain and the Fourier domain. The pixel domain constraints include the real-valued, boundedness, and color interpolation constraints and Fourier domain constraints [6].

Experimental Results with Estimated EFR: We compare the classification performances of the original EFR and the estimated EFR. Table 2 shows the confusion matrix for the estimated EFR using two-component ML-GMM. We notice that the classification accuracy with the estimated EFR is around 10% to 20% lower for certain manipulation categories compared with the corresponding results obtained with the original EFR reported in Table 1. Nevertheless, we can still differentiate categories using the estimated EFR with an accuracy close to 84.7%, suggesting that the estimation is effective.

We also examine the performance of the single component ML-GMM classifier, and notice that while the classification accuracies for (C2) to (C5) still remain the same, those for (C1) and (C6) reduce by around 10%. This suggests that (C1) and (C6) form loose clusters in terms of the estimated EFR. (C1) is loose since it is a collection of several different downsampling operations and each of them may have slightly different characteristics. We remark that operations grouped together based on intuitive signal processing understanding (our C1 to C6) may not actually be intrinsically similar. (C6) is loose due to the content dependence and the larger variation discussed in Section 3. With the imperfect estimates for EFRs, the large variation of (C6) incurs overlaps in the feature space, and thus lower the accuracy.

Multi-block Fusion: As discussed above, the EFR depends both on the camera and the image content. Such dependence is not desired since it lowers the inner-operation consistency. In this part, we introduce multi-block fusion as a possible approach to alleviate these problems. Assuming that the whole photograph or a certain significant portion of it undergoes the same operation, we can fuse evidence from more than one block to jointly determine the manipulation type. We adopt the naïve Bayes classifier which assumes that each block of the total N blocks is independent, and the a posteriori probability can be written as $P(C_i|F_1,\ldots,F_N) \propto \prod_{j=1}^{N} P(F_j|C_i)$, where C_i is the ith category, F_j is the estimated EFR of the jth block. Using the two-component GMM to model $P(F_j|C_i)$, our results show that multi-block fusion improves the classification accuracy from 84.7% to 93.0% with three blocks considered together.

5. CONCLUSIONS

In this paper, we introduce the Empirical Frequency Response (EFR) as a universal descriptor for digital tampering operations. We find that many LSI and non-LSI operations exhibit consistencies in the EFR, and therefore the EFR can be utilized to identify tampering operations when the input and output of the tampering module are known. Our results indicate that the proposed EFR based features can classify six categories of tampering with an accuracy of 95.9%. In scenarios where the system input is not available, we show that EFR can still be estimated just based on the output data, and used for tampering identification with an accuracy of 84.7%; which can be further improved to 93.0% by multi-block fusion. Experimental results supported by theoretical reasoning demonstrate the effectiveness of the proposed approach. Future work would include examining more types of tampering operations and modeling all factors that influence the EFR.

6. REFERENCES

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

[2] J. Lukas and J. Fridrich, "Estimation of primary quantization matrix in double compressed jpeg images," in *Proc. of the Digital Forensics Research Workshop*, Aug. 2003.

[3] H. Farid, "Blind inverse gamma correction," *IEEE Trans. on Image Proc.*, vol. 10, no. 10, pp. 1428–1433, Oct. 2001.

[4] I. Avcibas, S. Bayram, N. Memon, M. Ramkumar, and B. Sankur, "A classifier design for detecting image manipulations," in *Proc. of Inte. Conf. on Image Proc. (ICIP)*, Oct. 2004, vol. 4, pp. 2645–2648.

[5] H. Farid and S. Lyu, "Higher-order wavelet statistics and their application to digital forensics," in *IEEE Workshop on Statistical Analysis in Computer Vision*, June 2003.

[6] A. Swaminathan, M. Wu, and K. J. R. Liu, "Digital image forensics via intrinsic fingerprints," *IEEE Trans. on Infor. Forensics and Security*, vol. 3, no. 1, pp. 101–117, March 2008.

[7] T. S. Huang, Ed., *Two-Dimensional Diginal Signal Processing II: Transforms and Median Filters*, Springer-Verlag, 1981.