# FORENSIC ANALYSIS OF FULL-FRAME LINEARLY FILTERED JPEG IMAGES

*V. Conotter [#], P. Comesaña [#], F. Pérez-González [#⋆]*

[#] Signal Theory and Communications Department, University of Vigo, Spain
[⋆] Gradiant (Galician Research and Development Center in Advanced Telecommunications), Vigo, Spain

## ABSTRACT

The characteristic artifacts left in an image by JPEG compression are often exploited to gather information about the processing history of the content. However, linear image filtering, often applied as post-processing to the entire image (full-frame) for enhancement, may alter these forensically significant features, thus complicating the application of the related forensics techniques. In this paper, we study the combination of JPEG compression and full-frame linear filtering, analyzing their impact on the Discrete Cosine Transform (DCT) statistical properties of the image. We derive an accurate mathematical framework that allows to fully characterize the probabilistic distributions of the DCT coefficients of the quantized and filtered image. We then exploit this knowledge to estimate the applied filter. Experimental results show the effectiveness of the proposed method.

***Index Terms***— JPEG compression, linear filtering

## 1. INTRODUCTION

In today's digital age, low-cost and sophisticated digital technologies allow for the creation of sophisticated and visually compelling photographic fakes, thus questioning the trust in photographs. The nascent field of digital forensics has emerged to help regain some trust in digital photographs [1].

A plethora of forensic techniques have been proposed in the literature so far, aiming at identifying specific processing operators applied to images [2], but little attention has been paid to the forensic analysis of chains of operators. In such a scenario, difficulties in the authentication may arise since the characteristic footprints exploited to detect a specific processing may be wiped off by the application of a second one. JPEG compression is a forensically interesting processing operation, as it is one of the most popular image compression schemes in use today. Previous works in the literature exploit the characteristic artifacts introduced in the DCT distribution of an image during compression to discover instances of previous JPEG compression and even estimate the used quantization steps [3][4][5]. Multiple instances of JPEG compressions have been also studied, detecting instances of all previous compressions [6]. However, it is very likely that later in its life the JPEG-compressed image will be processed by a further full-frame post-processing aiming at reducing the JPEG-compression block effects, or at enhancing the quality of the image. By doing so, the characteristic artifacts present in the DCT distribution of JPEG images may be partially perturbed, thus complicating the application of the given forensic techniques (see [7], where authors study the case when image resizing has been applied between two compressions).

In this work, we study the combination of single JPEG compression and full-frame linear post-processing. Specifically, linear filtering represents an interesting case study since it is a very common and useful linear tool applied for image enhancement, such as edge sharpening, noise removal, illumination correcting and deblurring. We analyze the DCT statistical properties of a JPEG compressed and linearly-filtered image and mathematically establish the relationship between DCT coefficients, before and after filtering. By exploiting the well-known statistical properties of the distribution of DCT coefficients in JPEG images, we derive a model to theoretically characterize the probability distribution of the DCT coefficients of a JPEG image filtered with a given filter kernel. As a first result of this analysis, we show that the extended assumption for natural images about the AC DCT coefficients for different frequencies being independent, and for the same frequency being i.i.d. [8][9], does not hold; indeed, the inter- and intra-block redundancy of the quantized DCT coefficients must be taken into account. By considering those redundancies, the studied processing can be accurately modeled. In order to finally identify the applied linear filter, a distinguishability measure is calculated to quantify the difference between the derived models (each model depends on the applied filter kernel) and the actual distribution of a to-be-tested image. The minimum difference will be taken as evidence for the identification of the applied filter. We will assume in this work the quantization to be fixed and known, although future work will be devoted to removing this assumption.

To the best of our knowledge, the presented work constitutes a first attempt to study the statistical perturbation introduced by linear filters on JPEG images and, although their detection is not necessarily proof of malicious tampering, the derived framework represents a valuable mean to disclose the processing history.

## 2. BLOCK-WISE JPEG COMPRESSION AND FULL-FRAME LINEAR FILTERING

Throughout this paper, lower case letters (e.g., $x$) denote $L_1 \times L_2$ size images in the spatial domain;[1] $x(i,j)$ stands for the pixel at position $(i,j)$ of image $x$, with $i \in \{0,\dots,L_1-1\}$ and $j \in \{0,\dots,L_2-1\}$. Images in the $8\times 8$-DCT domain are denoted with uppercase letters (e.g., $X$), so $X^{i_8,j_8}(i',j')$ stands for the $(i',j')$ DCT coefficient at the $(i_8,j_8)$ block, where $i',j' \in \{0,\dots,7\}$, $i_8 \in \{0,\dots,(L_1/8)-1\}$, and $j_8 \in \{0,\dots,(L_2/8)-1\}$. For the sake of notational simplicity, and due to the similarity with the pixel domain notation, we will also use $X(i,j)$, where $i = i' + 8i_8$ and $j = j' + 8j_8$, to denote $X^{i_8,j_8}(i',j')$. Consequently, prime variables will denote modulo 8 reduced variables, e.g., $i' = i \bmod 8$.

Research supported by the European Union under project REWIND (Grant Agreement Number 268478), the European Regional Development Fund (ERDF) and the Spanish Government under projects DYNACS (TEC2010-21245-C02-02/TCM) and COMONSENS (CONSOLIDER-INGENIO 2010 CSD2008-00010), and the Galician Regional Government under projects "Consolidation of Research Units" 2009/62, 2010/85 and SCALLOPS (10PXIB322231PR).

[1] For the sake of simplicity, we will assume $L_1$ and $L_2$ to be integer multiples of 8.

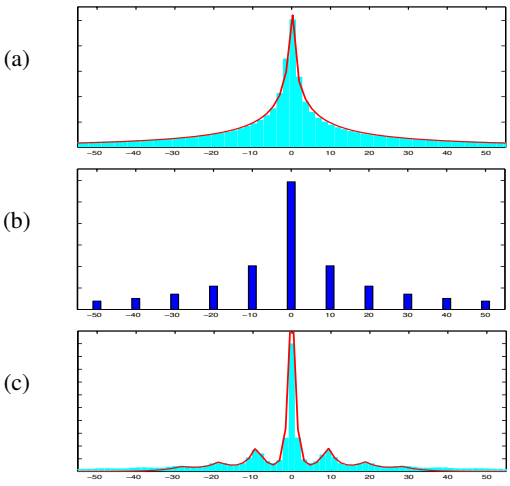


**Fig. 1**: Panel (a) shows the original DCT histogram for frequency $(0, 1)$ for uncompressed images and its curve fitting. In panel (b), the same distribution after quantization with step $\Delta(0, 1) = 10$ is presented. Panel (c) shows the given distribution when a linear average filter of size $3 \times 3$ has been further applied, together with the derived model for such distribution (red line).

The JPEG standard provides a block-based compression scheme, which operates on $8\times8$ non-overlapping blocks of DCT coefficients. Each DCT coefficient is quantized by dividing it by the corresponding entry $\Delta(i', j')$ of a quantization table, fixed based on the desired compression quality. Such quantized coefficients are then entropy-encoded (Huffman coding) and stored in the JPEG file format. The compressed data stream can be decompressed, applying all the steps in reverse order. The compression process forces the values of the DCT coefficients to be clustered around multiples of the quantization step $\Delta(i', j')$. These specific artifacts in the coefficient distribution can be shown in Fig. 1. Panel (a) shows the histogram of the DCT coefficients at frequency $(0, 1)$ collected from un-compressed images [10], while panel (b) depicts the distribution of the same data after quantization, with $\Delta(0, 1) = 10$. It becomes clear that the structure of such histogram is related to the employed quantization step. For the sake of presentation we disregard the round-off and truncation errors, in the pixel domain, introduced by the compression scheme, without affecting the conducted analysis. However, linear image processing, often applied as post- processing for image enhancement, may alter the characteristic statistical properties introduced by the compression scheme. Here, we study the case of linear filtering since it is a very common and powerful tool employed for image enhancement. It operates by convolving the original image with an appropriate filter kernel. The result of such a convolution is a filtered image, whose pixel values are a weighted sum of a certain number of neighboring pixels, depending on the filter size. Fig. 1(c) shows the histogram of the DCT frequency coefficients of panel (b) after filtering with an $3 \times 3$ Average filter; the characteristics of the histogram of the quantized coefficients are clearly perturbed, but new patterns appear, depending both on the employed quantization factor and the filter kernel. In this work we study such artifacts in order to identify the filter kernel a JPEG image has undergone. We mathematically analyze the statistical properties of the DCT distribution of compressed and filtered images and derive an accurate model for them. Fig. 1(c) serves as an example of the derived model for the probability distribution of JPEG filtered images.

## 3. MATHEMATICAL MODEL

In order to derive the theoretical model characterizing the DCT distribution of a JPEG and filtered image, we firstly mathematically express the deterministic relation between the quantized DCT coefficient $X(i, j)$ and that of the JPEG and filtered image $Y(i, j)$. The study case here concerned is shown in Fig. 2. Following the scheme backwards, we start considering the DCT coefficients $Y(i, j)$ of a JPEG compressed and filtered image and transform them into the spatial domain $y(i, j)$. Through a linear convolution operation with the filter kernel $\mathbf{h}$, we can relate $Y(i, j)$ to the pixels of the compressed image $x$, which is the Inverse DCT of $X(i, j)$. Exploiting the linearity property of both the filtering operation and the DCT transform and working out the math, we are finally able to mathematically formulate the relation between the DCT coefficients of a quantized image $X(\cdot, \cdot)$ and the DCT coefficients of the further filtered image $Y(\cdot, \cdot)$. Specifically, given a filter kernel of size smaller than or equal to 17, the coefficients $X(i, j)$ contributing in the calculation of $Y(i, j)$ are those from the same block of $Y(i, j)$ plus those from the 8 immediate surrounding blocks, resulting in $24 \times 24$ coefficients. It becomes clear that for filter kernels of sizes larger than 17, the number of contributing coefficients would increase, involving more surrounding blocks. Finally, we make explicit the contribution of the DCT coefficient $X(i, j)$ at the same position of $Y(i, j)$, as follows:

$$Y^{i_8,j_8}(i', j') \quad = \quad \gamma_{i',j'} \cdot X^{i_8,j_8}(i', j') + N^{i_8,j_8}(i', j'), \quad (1)$$

where $\gamma_{i',j'}, N^{i_8,j_8}(i', j') \in \mathbb{R}$ are a frequency dependent scaling factor and noise term, respectively. These two terms can be calculated, through some math (not reported here because of space constraints), according to (2), where $\text{DCT}_{i',j'}$ is the $(i', j')$-th DCT coefficient obtained from an $8 \times 8$ pixel block, $\text{IDCT}_{i',j'}$ is the $8 \times 8$ pixel block (located at $\{8i_8, \ldots, 8i_8 + 7\} \times \{8j_8, \ldots, 8j_8 + 7\}$) obtained by applying the IDCT to the $(i, j)$ DCT coefficient ($i_8 = \lfloor i/8 \rfloor$, $j_8 = \lfloor j/8 \rfloor$), $*$ denotes the bidimensional convolution, and $[\mathbf{A}]_{a,b}^{c,d}$ denotes the submatrix of $\mathbf{A}$ with first index taking values in $\{a, \ldots, b\}$, and second index in $\{c, \ldots, d\}$. $N^{i_8,j_8}(i', j')$ stands for the second term in the summation in (2).

Given the derived deterministic expression in (1) for $Y(i, j)$, we can exploit the knowledge about the distribution of the quantized coefficients $X(i, j)$ to study the distribution of the DCT coefficients of the final image $y$. Usually, the probability distribution of DCT coefficients in natural images is modeled as a zero-mean Generalized Gaussian Distribution (GGD) [3]. Due to quantization, the probability distribution of each quantized DCT coefficient is [3]:

$$f(X^{i_8,j_8}(i', j') = k\Delta|\Delta) = \int_{(k-\frac{1}{2})\Delta}^{(k+\frac{1}{2})\Delta} f_{GGD_{i',j'}}(\tau)d\tau, \quad (3)$$

where $k \in \mathbb{Z}$ and, for the sake of notation simplicity, $\Delta = \Delta(i', j')$. It becomes clear that the probability mass function of each frequency coefficient of a JPEG image (Fig. 1(b)) presents specific artifacts, whose structure is related to the quantization step. From probability theory [11], given two discrete independent random variables, the probability density function (pdf) of their sum is the convolution of their corresponding pdfs. Therefore, according to the derived mathematical model in (1), and based on the common DCT coefficients models, which typically assume the different frequency components to be independent and the coefficients in a given frequency to be i.i.d. [12], we would expect the probability distribution of the DCT coefficient $Y(i, j)$ to be the result of a convolution between a train of impulses located at $\gamma_{i',j'} \cdot k\Delta(i', j')$, with $\gamma \in \mathbb{R}$, and a noise component due to the contributions of all the neighbor-

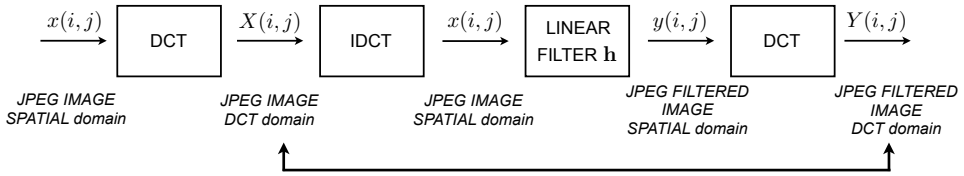


**Fig. 2**: Block scheme of the considered processing operations. In order to exploit the known statistical properties of the distribution of $X(i,j)$, we establish its mathematical relationship with the DCT coefficient $Y(i,j)$ of the quantized and filtered image.

$$Y(i,j) = \mathrm{DCT}_{i',j'}\left(\left[\mathbf{h} * \mathrm{IDCT}_{i',j'}\left(X(i,j)\right)\right]_{\lfloor \frac{i}{8} \rfloor, \lfloor \frac{j}{8} \rfloor}^{\lfloor \frac{i}{8} \rfloor + 7, \lfloor \frac{j}{8} \rfloor + 7}\right) + \sum_{(k_1,k_2)\in\{-8,\ldots,15\}^2,(k_1,k_2)\neq(i',j')} \mathrm{DCT}_{i',j'}\left(\left[\mathbf{h} * \mathrm{IDCT}_{k_1',k_2'}\left(X\left(\left\lfloor \frac{i}{8} \right\rfloor + k_1, \left\lfloor \frac{j}{8} \right\rfloor + k_2\right)\right)\right]_{\lfloor \frac{i}{8} \rfloor - \lfloor \frac{k_1}{8} \rfloor, \lfloor \frac{j}{8} \rfloor - \lfloor \frac{k_2}{8} \rfloor}^{\lfloor \frac{i}{8} \rfloor - \lfloor \frac{k_1}{8} \rfloor + 7, \lfloor \frac{j}{8} \rfloor - \lfloor \frac{k_2}{8} \rfloor + 7}\right). \quad (2)$$

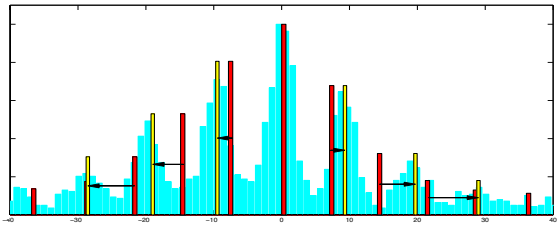

**Fig. 3**: Probability distribution, at frequency $(0,1)$, quantized with a step $\Delta(0,1) = 10$ and filtered with averaging filter. The red impulses represent the location of $\gamma_{0,1} \cdot k\Delta(0,1)$ while the yellow ones are translated by the mean of the Noise component.

ing coefficients, according to (1). However, we show that indeed the typical assumptions on the DCT coefficients distribution of natural images do not hold, thus resulting in a deviation between the classical theoretical models and the empirical data. Specifically, a scaling between the peaks in the histogram of the DCT coefficients of the compressed and filtered image and the impulse train identifying the location of the translated quantization step $\gamma_{i',j'} \cdot k\Delta(i',j')$ is observed. This suggests the need for a different model for the noise component in (1), which cannot any longer be considered as the addition of independent variables (coefficients of different frequencies) and i.i.d. components (coefficients in the same frequency). We model the noise component as a GGD variable, with parameters depending on the centroid $\gamma_{i',j'} \cdot k\Delta(i',j')$ the noise is centered around. We analyze the mean of such noise and verify that for real images it monotonically increases with the quantized samples value. In Fig. 3 it is shown how the translated impulses, now centered at $\gamma_{i',j'} \cdot k\Delta(i',j')$ plus the mean of the noise component, match the peaks of the histogram. Taking into account the scaling inferred by the noise component, an accurate model for the distribution of the DCT coefficients of a filtered JPEG image can be derived.

## 4. PROPOSED FORENSIC APPROACH

The main goal of the proposed approach is to estimate the filter operator an image has been gone through. We build a dictionary-based database for the derived models corresponding to different applied filters, and a distinguishability measure is calculated to quantify the difference between the theoretically derived models and the actual distribution of an image. So, in our scheme, this measure will be taken as an evidence to identify the filter operator applied to the under-test image. As a first attempt in this forensic case study, we assume the quantization applied to the image during compression to be known a priori. Future work will release this assumption.

To build a generalized model associated with each filter in the dictionary, we proceed as follows:

- We collect the DCT frequency coefficients from 669 of the images present in the UCID- Uncompressed Image Database [10]. We then fit, for each frequency, a GGD (see Fig. 1(a)).
- We evaluate the distribution of the quantized coefficients, according to (3) (see Fig. 1(b)). In this work we used compression quality factors $QF \in \{40, 50, 60, 70, 80, 90\}$.
- We select a set of linear filters to be part of the dictionary, both low-pass (LP) and high-pass (HP) (e.g., Moving Average, Gaussian, Laplacian), with different settings for the window size, the variance $\sigma^2$ or the scale parameter $\alpha$, as reported in Table 1.
- For each filter and AC DCT coefficient, we calculate $\gamma_{i',j'}$ in (1), according to (2).
- For each $\gamma_{i',j'} \cdot k\Delta(i',j')$ the corresponding noise component is modeled as a GGD.
- The distribution of DCT coefficients, quantized and filtered with a given kernel, will be the sum of GGDs, each of them centered at $\gamma \cdot k\Delta(i',j')$ translated by the mean of the noise component and with amplitude depending on the distribution of the quantized and not filtered coefficients (e.g., see Fig. 1(c)).

The distinguishability among the derived models for the considered filters is calculated by means of the $\chi^2$ histogram distance [13]:

$$\chi^2 = \frac{1}{2}\sum_z \frac{(P_z - Q_z)^2}{(P_z + Q_z)}, \quad (4)$$

where $P$ and $Q$ are the two probability distributions to be compared, which of course depend on the considered DCT frequency. Intuitively, the $\chi^2$ distance would tend to zero when one distribution converges to the other. So, a lower $\chi^2$ measure will be the clue to identify the filter operator applied to the under-test image.

We combine the independently calculated comparisons for each analyzed frequency coefficient by summing their values, according to the Minkowski norm [14]. As a preliminary study we selected a subset of frequency coefficients to be representative of low, medium and high frequencies and identified the most significative ones. As a future work, we plan to further explore the effect of all the frequency coefficients. As a first result, we observed a good distinguishability among all the theoretical models, even if very low $\chi^2$ values may mislead the correct classification of some of the filter, as in the case of filters (1)-(2)-(4)-(6) (see Table 1). This is due to the similarity of the frequency response of those filters for the analyzed DCT coefficients, and consequently, of their corresponding models. This issue is not specific of the presented framework, but a general constraint. Table 1 shows groups based on the similarity

| | |
|---|---|
| 1. LP Average $[3 \times 3]$ | 3. LP Gaussian $[3 \times 3]$, $\sigma^2$= 0.5 |
| 2. LP Average $[5 \times 5]$ | 5. LP Gaussian $[5 \times 5]$, $\sigma^2$= 0.5 |
| 4. LP Gaussian $[3 \times 3]$, $\sigma^2$= 1 | 7. HP Laplacian, $\alpha^2$= 0.2 |
| 6. LP Gaussian $[5 \times 5]$, $\sigma^2$= 1 | 8. HP Laplacian, $\alpha^2$= 0.7 |
| 9. LP Laplacian, $\alpha^2$= 0.2 | 11. HP Average $[3 \times 3]$ |
| 10. LP Laplacian, $\alpha^2$= 0.7 | 12. HP Average $[5 \times 5]$ |

**Table 1**: Filters present in the dictionary-based filter database, grouped according to the similarity of their frequency response.

of the frequency response of the filters selected to be part of the dictionary. Given the models for all the distributions corresponding to a specific filter, the performance of the proposed algorithm is verified in terms of percentage of correct classification over the image database. Each of the 669 images in the database [10], not previously used to build the models, is compressed with quality factors QF $\in \{40, 50, 60, 70, 80, 90\}$ and post-processed with each of the filter kernels present in the dictionary. We then compare the obtained DCT histogram, for each frequency, with all the corresponding DCT coefficient pdfs derived in the steps described above. The estimated applied filter is that providing the minimum $\chi^2$ distance. Accuracy of correct classification is reported in Fig. 4(a). Results are reported as grouped bar graphs, where each group refers to a fixed quality factor QF and each bar represents one of the 13 filters in the dictionary. Results are promising, even if accuracy is low for some filters with high QF. The average accuracy, over all the groups, is reported as a black line in Fig. 4(c). We notice that as the QF increases, the average accuracy decreases. This is due to the fact that with a high QF, the compression steps $\Delta(i', j')$ for low frequencies are small and thus difficult to be localized after filtering, while higher frequencies will be likely set to zero, especially for LP filtering. For each QF, we identified, among all the selected frequency coefficients, a set of them to be the best combination which better help to identify the applied filter in terms of the highest accuracy. For example, we have the sets $\{(0,1),(1,0),(5,2)\}$ for QF $= 40$ and set $\{(0,1),(4,3),(5,2),(5,3),(5,0)\}$ for QF$= 90$. Due to space constraints, coefficients for different QFs are not reported. We note that lower frequencies are more important for lower QFs, since it is likely that higher frequencies will be set to zero.

Intuitively, lower frequencies will also be more significant when dealing with low-pass filters, while higher frequencies will be needed to correctly identify high-pass filters. Based on this idea, we divide the filters database into two halves, one with low-pass filters, while in the second are only high-pass filters. Depending on the type of the filter applied to the under-test image, the proposed algorithm will be applied using one of the two parts of the database. To automatically determine if an image has been low-pass or high-pass filtered, we build a set of reference Fourier Transform images, each one corresponding to a filter in the set of low/high-pass kernels and a particular compression factor QF. In order to do that, we average 100 FFTs of images compressed with the selected QFs and filtered with the given filter; minimizing the Mean Square Error between the reference FFTs and that of the under-test image we are able to easily classify the image, with a $100\%$ accuracy. So, using the LP and the HP filters database separately, we obtained accuracy results as reported in Fig. 4(b), where high-pass filters are grouped with the red color, while low-pass are depicted with blue. An improved percentage of correct classification is achieved for each set of filters, compared to the classification with all the filters together (Fig. 4(a)). Fig. 4(c) reports the average accuracy for the two separated classification, considering only low-pass (red line) or high-pass filters

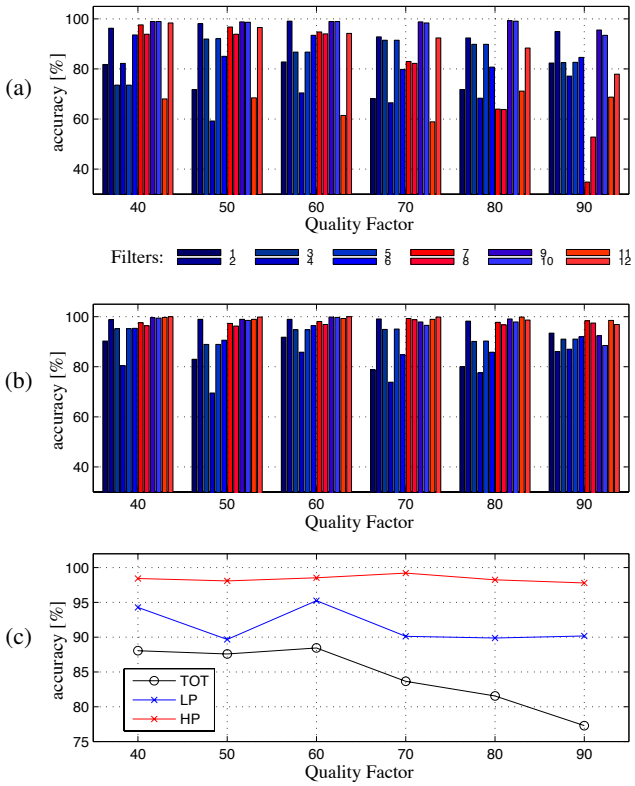


**Fig. 4**: Panel (a) shows the classification accuracy when each of the 13 filters in the dictionary is applied to each image in the database, compressed with different QFs, while panel (b) shows the accuracy when classifying LP (blue) and HP (red) filters separately. Panel (c) shows the average accuracy for the classification of all the filters, only LP and only HP filters (black,blue and red line, respectively).

(blue line). We notice that an improved accuracy is achieved with the separation of the filters. Again, we identified specific sets of coefficients to be utilized in the separated classifications. As an example, for QF $= 40$ we have the set $\{(0,1),(1,0),(1,1),(2,6)\}$ for LP filters, and $\{(7,0),(5,0),(0,7)\}$ for HP filters. So, for HP filters higher frequencies are more significant, as expected. Similar sets of coefficients are found for higher QFs. The presented results are very promising and show the efficacy of the proposed technique.

## 5. CONCLUSION

We have presented a mathematical model to characterize the DCT coefficients distributions of a full-frame linearly-filtered JPEG image. We explicitly expressed the theoretical relationship between the DCT coefficients before and after filtering so as to accurately analyze the effect of the considered processing. The derived theoretical model allows building a dictionary-based database of distributions of quantized images being filtered with a given kernel. We exploited such dictionary for estimating the filter given the quantization, by using the $\chi^2$ distance as target function. The presented framework represents a first attempt to analyze the effects of full-frame linear filtering on block-based compressed images. Future work will be devoted to enlarge the filter dictionary and to eliminate the assumption on the knowledge of the quantization step, so that eventually this framework may be regarded as a forensically helpful means to jointly disclose the applied compression factors and the filter kernel.

## 6. REFERENCES


[1] E. Delp and M. Wu, "Digital forensics," *IEEE Signal Processing Magazine*, vol. 2, no. 26, pp. 14–15, 2009.

[2] J. Rhedi, W. Taktak, and J-L. Dugelay, "Digital image forensics: a booklet for beginners.," *Multimedia Application Tools*, vol. 51, pp. 133–162, 2011.

[3] R. Neelamani, R. de Queiroz, Z. Fan, S. Dash, and R. Baraniuk, "JPEG compression history estimation for color images," *IEEE Transactions on Image Processing*, vol. 15, no. 6, pp. 1365–1378, 2006.

[4] W. Luo, J. Huang, and G. Qiu;, "JPEG error analysis and its application to digital image forensics," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 3, pp. 480–491, 2010.

[5] D. Fu, Y. Shi, and W. Su, "A generalized benford's law for JPEG coefficients and its applications in image forensics," *SPIE Conference on Security, Steganography, and Watermarking of Multimedia Contents*, vol. 6505, 2007.

[6] T. Pevný and J. Fridrich, "Detection of double-compression in JPEG images for applications in steganography," *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 2, pp. 247–258, 2008.

[7] T. Bianchi and A. Piva, "Reverse engineering of double JPEG compression in the presence of image resizing," *IEEE International Workshop on Information Forensics and Security (WIFS)*, pp. 127–132, 2012.

[8] I. J. Cox, *Digital Watermarking and Steganography*, Morgan Kaufmann Publishers, 2008.

[9] M. Barni and F. Bartolini, *Watermarking systems engineering - Enabling digital assets security and other applications*, CRC Press, 2004.

[10] G. Schaefer and M. Stich, "UCID - an uncompressed colour image database," *SPIE Conference on Storage and Retrieval Methods and Applications for Multimedia*, pp. 472–480, 2004.

[11] A. Papoulis, *Probability, Random Variables and Stochastic Processes*, McGraw-Hill International Edition, 1991.

[12] P. Moulin and M. Kvanç Mihçak, "A framework for evaluating the data-hiding capacity of image sources," *IEEE Transactions on Image Processing*, vol. 11, no. 9, pp. 1029–1042, 2002.

[13] O. Pele and M. Werman, "The quadratic-chi histogram distance family," *Computer Vision - ECCV 2010*, vol. 6321, pp. 749–762, 2010.

[14] J. Puzicha, T. Hofmann, and J. Buhmann, "Non- parametric similarity measures for unsupervised texture segmentation and image retrieval," *IEEE Computer Society Conference on Computer Vision and Patter Recognition (CVPR)*, pp. 267–272, 1997.