JPEG Re-Compression Detection

Xiaoying Feng and Gwenaël Doërr

Department of Computer Science – University College London
Malet Place – London WC1E 6BT – United Kingdom

ABSTRACT

Re-quantization commonly occurs when digital multimedia content is being tampered with. Detecting re-quantization is therefore an important element for assessing the authenticity of digital multimedia content. In this paper, we introduce three features based on the observation that re-quantization (i) induces periodic artifacts and (ii) introduces discontinuities in the signal histogram. After validating the discriminative potential of these features with synthetic signals, we propose a system to detect JPEG re-compression. Both linear (FLD) and non-linear (SVM) classifications are investigated. Experimental results clearly demonstrate the ability of the proposed features to detect JPEG re-compression, as well as their competitiveness compared to prior approaches to achieve the same goal.

Keywords: Multimedia forensics, Re-quantization

1. INTRODUCTION

In the last 60 years we have witnessed an astonishing revolution in the way information is presented and handled. Starting from a situation in which information was basically represented by analogue signals that could be directly perceived by human senses, scientific and technological advancements have made possible a shift toward an abstract representation of information, whereby information is reduced to a sequence of bits. This shift, first slowly, then more and more rapidly, has led to the current digital age where most information is created, captured, transmitted, stored and processed in digital form. Although representing information in digital form has a number of evident advantages, it also raises new issues and challenges. In particular, the ease with which digital contents can be manipulated casts serious doubts on their validity as trustworthy representation of reality.[1] These doubts are strengthened with each new report of tampered content, and as doubts and skepticism increase, our trust in what we see and what we hear is eroded. Today more than ever, our perception of reality might be misled by the limitations of our senses.[2] The consequences are enormous and touch every aspects of our life. Did a politician utter such a sentence? Is a war picture provided by a photo-reporter a faithful representation of what is happening on the field? Did an offensive act captured with a video camera actually occur? At a more general level, how can we build and reinforce our beliefs if nothing of what we can see or hear can be trusted? There is no opportunity for rational debate if no evidence is trusted. As a result, there is today a crucial need to restore and maintain trust in our primary sources of information.

Over the last few years, the body of scientific techniques for recovering evidence from digital data has been growing steadily. Digital forensics is commonly defined as "the use of scientifically derived and proven methods toward the preservation, collection, validation, identification, analysis, interpretation, documentation and presentation of digital evidence derived from digital sources for the purpose of facilitating or furthering the reconstruction of events found to be criminal, or helping to anticipate unauthorized actions shown to be disruptive to planned operations".[3] Multimedia forensics focus on those aspects of digital forensics concerned with multimedia signals, e.g. still images, videos, audio clips, etc. There are mainly two approaches to multimedia forensics. The first one, referred to as *active forensics*, relies on modifying the multimedia signal prior to its distribution to assist later forensic analysis. Typical examples of active forensics include digital watermarks and cryptographic signatures. However, a limitation of active forensics technologies is the need for them to be embedded in all content-generating devices (sensors, microphones, cameras, ...), which is only possible in very

Further author information (Send correspondence to G.D.) X.F.: E-mail: x.feng@cs.ucl.ac.uk, Telephone: +44 (0)20 7679 0349; G.D.: E-mail: g.doerr@cs.ucl.ac.uk, Telephone: +44 (0)20 7679 3570.

restricted cases in practice. In many situations, there is no opportunity to employ active technologies. We are simply presented with a signal and we must decide its authenticity based on the analysis of the signal itself. *Passive forensics* techniques[4] concentrate on answering three kinds of questions: (i) which device has been used to capture this multimedia content (proof of origin), (ii) has this multimedia content been tampered with (tamper detection), and (iii) which signal processing primitives have been applied to this multimedia content (tell-tale forensics).

In this paper, we will focus on one specific multimedia forensics technique, that is to say re-quantization detection. Re-quantization usually occurs when multimedia content is successively compressed with different parameters. For instance, steganographic algorithms such as F5[5] or Outguess,[6] by design, recompress the processed JPEG images. In this case, re-quantization detection is known to provide valuable information in order to improve steganalysis.[7] Re-quantization also often happens when digital images are tampered with. Indeed, still images usually have to be decompressed in order to be edited, hence potentially inducing re-quantization when the manipulated image is later re-compressed for storage. Re-quantization detection can thus be exploited to assess whether an image has been tampered with or not.[8] Similarly, re-quantization detection could be useful to improve the security of conditional access systems relying on speaker verification. In this scenario, a key question is: does the input audio signal originate from a live person or is it a recording of the person? Re-quantization artifacts are likely to be present in the latter case. In Section 2, we describe the process of re-quantization and review relevant prior works in the area. Subsequently, we introduce in Section 3 three features extracted from the approximation error histogram and demonstrate their discriminative performances on synthetic signals. We then focus on the particular problem of JPEG re-compression detection in Section 4 and report detection performances compared to two state-of-the-art baseline systems. Finally, Section 5 summarizes the contributions of the paper and highlights a couple of issues remaining for future work.

2. PRIOR WORKS ON UNIFORM RE-QUANTIZATION

Let us consider a generic 1-D signal $\mathbf{x}(t)$ for illustrative purpose. Uniform quantization is mathematically described by the following one-parameter function:

$$\mathrm{Q}_{\Delta}\left(\mathbf{x}(t)\right) = \Delta \times \mathrm{q}_{\Delta}\left(\mathbf{x}(t)\right), \tag{1}$$

where $\Delta \in \mathbb{R}_+^*$ is the quantization step size and $\mathrm{q}_\Delta(.)$ is a quantization index function, which assigns to each sample $\mathbf{x}(t)$ an integer value $\mathrm{q}_\Delta(\mathbf{x}(t)) \in \mathbb{Z}$. There is no standard for computing the quantization index and several alternate definition are used in practice, e.g.:

$$\mathrm{q}_\Delta(\mathbf{x}(t)) = \left\lfloor \frac{\mathbf{x}(t)}{\Delta} \right\rfloor, \tag{2a}$$

$$\mathrm{q}_\Delta(\mathbf{x}(t)) = \left\lfloor \frac{\mathbf{x}(t)}{\Delta} + 0.5 \right\rfloor, \tag{2b}$$

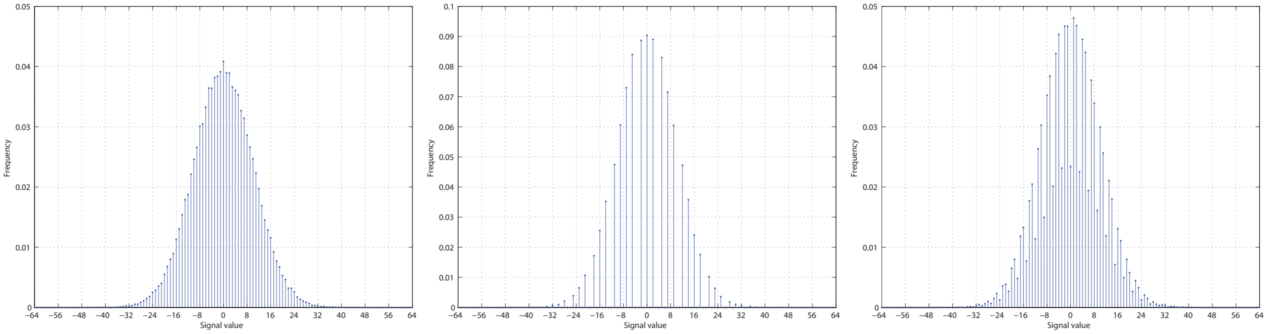
$$\mathrm{q}_\Delta(\mathbf{x}(t)) = \mathrm{sign}(\mathbf{x}(t)) \times \left\lfloor \frac{|\mathbf{x}(t)|}{\Delta} + 0.5 \right\rfloor. \tag{2c}$$

For simplicity, we will only consider Equation (2c) in the remainder of this article as it corresponds to the default rounding operation in Matlab. Re-quantization then reduces to a succession of two uniform quantization operations with quantization step sizes Δ_1 and Δ_2:

$$\mathbf{z}(t) = \mathrm{Q}_{\Delta_2}\left(\mathbf{y}(t)\right) = \mathrm{Q}_{\Delta_2}\left(\mathrm{Q}_{\Delta_1}\left(\mathbf{x}\left(t\right)\right)\right), \tag{3}$$

where $\mathbf{y}(t)$ and $\mathbf{z}(t)$ are respectively the single quantized signal and the re-quantized signal. The re-quantization ratio is defined as $\rho = \Delta_1/\Delta_2$.

Figure 1 illustrates typical artifacts introduced by the re-quantization process in the histogram of the tested signal. The original signal follows a Gaussian distribution and has been quantized using different quantization step sizes. As highlighted in earlier works,[9,10] the nature of requantization artifacts differ significantly depending on whether the requantization ratio ρ is smaller or larger than 1.



(a) Single quantization

(b) Re-quantization with $\rho = 2.3$

(c) Re-quantization with $\rho = 0.6$

Figure 1. Illustration of the artifacts introduced by re-quantization in the histogram of a normally distributed signal for different re-quantization ratios. For single quantization, the step size has been set equal to 1; for re-quantization, the second step size is also set to 1.

When the second quantizer is finer than the first one ($\rho > 1$ as in Subfigure 1(b)), the re-quantization process periodically introduces empty bins in the histogram due to the fact that there are fewer bins after the first quantization than that after the second one. Note the particular case when the reconstruction points of the two quantizers are aligned i.e. $\rho \in \mathbb{N}^*$. The histogram of the requantized signal $\mathbf{h_z}(k)$ can then be written as:

$$\mathbf{h_z}(k) = \int_{k\Delta_2 - \frac{\Delta_1}{2}}^{k\Delta_2 + \frac{\Delta_1}{2}} \mathrm{pdf}_{\mathbf{x}}(u).\mathrm{d}u \times \sum_{i=-\infty}^{+\infty} \delta_{k\Delta_2, i\Delta_1} = \mathrm{I}_{\mathbf{x}}(k\Delta_2) \times \mathrm{D}_{\Delta_1}(k\Delta_2), \quad k \in \mathbb{Z}, \tag{4}$$

where $\mathrm{pdf}_{\mathbf{x}}()$ is the probability density function of the original signal $\mathbf{x}(t)$ and $\delta_{i,j}$ is the Kronecker delta function. The histogram is the multiplication of some function $\mathrm{I}_{\mathbf{x}}(.)$ with a Dirac comb of periodicity Δ_1. As a result, the Fourier transform of the histogram is the convolution of the Fourier transform of $\mathrm{I}_{\mathbf{x}}(.)$ with a Dirac comb of periodicity $1/\Delta_1$. In other words, the Fourier transform of $\mathrm{I}_{\mathbf{x}}(.)$ is replicated at different locations, hence introducing extra peaks in the frequency domain as mentioned in previous papers.[9] Although derivations are not manageable when $\rho \notin \mathbb{N}^*$, the re-quantization process still introduces extra peaks in the Fourier transform.

In contrast, when the second quantizer is coarser than the first one ($\rho < 1$ as in Subfigure 1(c)), the number of bins after the first quantization is larger than after the second quantization. As a result, several bins are merged together during the second quantization and if the final bins do not receive exactly the same number of bins from the first quantizer, it results in strong discontinuities which indicate re-quantization. The smaller the re-quantization ratio becomes, the less noticeable are the discontinuities. It should be noted that when $1/\rho \in \mathbb{N}^*$, during the second quantization, each bin receives the same number of bins and thus, no discontinuity is created. In other words, re-quantization is undetectable. This limitation of re-quantization detection was already reported in.[11]

Despite the fact that these artifacts have now been identified for a long time, there is no definitive algorithm that will simply report whether some signal is re-quantized or not. In early research, detection relied on the visual inspection of the magnitude of the spectrum of the histogram to isolate unexpected peaks.[9, 10] In parallel, automatic techniques have been developed relying on powerful machine learning techniques such as neural networks[11] or Support Vector Machine (SVM)[12] applied to the sample histogram itself. Although such techniques, which exploit 'black-box' machine learning, have the potential to achieve reasonable performances, they fail to isolate the few discriminating features characterizing re-quantization. A first attempt in this direction has been to measure how well the first digit of the DCT coefficients in a JPEG image respects a generalized Benford's law,[13] since re-quantization seems to make the histogram deviate from this model. However, recent works reported that classification performances were close to random guessing with a bias toward the 'single quantized' class. This research has been subsequently extended to detect localized traces of re-quantization.[8]

The strategy consists of modeling the global histogram as a composite of two histograms, one single quantized and the other re-quantized, and to locally decide to which class the pixel/block belongs based on a Bayesian probabilistic model.

3. DISCRIMINATIVE FEATURES EXTRACTION FOR CLASSIFICATION

In this Section, we first describe 3 features derived from the signal histogram and then demonstrate the potential to achieve accurate discrimination between single quantization and re-quantization on synthetic signals.

3.1 Description of the Features

Early work has pointed out that re-quantization introduces extra peaks in the magnitude of the Fourier transform of the histogram. In an attempt to capture this characteristic, a first feature $\mathbf{f}_1$ is derived from the Fourier transform $\mathbf{H}(u)$ of the histogram $\mathbf{h}(k)$ of the tested signal. The idea consists of computing the average magnitude of the spectrum at the loci of local maxima as follows:

$$\begin{aligned} \mathbf{f}_1 &= \frac{1}{|\mathcal{U}|}\sum_{u\in\mathcal{U}} \mathbf{M}(u), \\ & \mathcal{U} = \Big\{u : \big(\mathbf{M}(u) - \mathbf{M}(u-1)\big)\big(\mathbf{M}(u) - \mathbf{M}(u+1)\big) > 0 \;\wedge\; \big(\mathbf{M}(u) - \mathbf{M}(u+1)\big) > 0\Big\}, \end{aligned} \tag{5}$$

where $\mathbf{M}(u) = |\mathbf{H}(u)|$ is the magnitude of the histogram spectrum. Special care is taken to remove the DC component index from $\mathcal{U}$ and to manage borders. Since re-quantization introduces extra-peaks in the histogram spectrum, the value of this feature is expected to be larger, on average, for re-quantized signals than for single quantized ones.

When surveying the different histograms depicted in Figure 1, one could argue that re-quantization is actually revealed by the presence of discontinuities in the histogram $\mathbf{h}(k)$. In order to look for the presence of such discontinuities, we look at the approximation error $\mathbf{d}(k)$ defined below:

$$\mathbf{d}(k) = \mathbf{h}(k) - \frac{\mathbf{h}(k-1) + \mathbf{h}(k+1)}{2}. \tag{6}$$

If the tested signal is single quantized, its histogram is smooth and the approximation error is expected to be close to zero. Conversely, when the tested signal is re-quantized, strong discontinuities are observed in the histogram which result in large values in the approximation error. In order to differentiate between the two situations, a second feature is introduced:

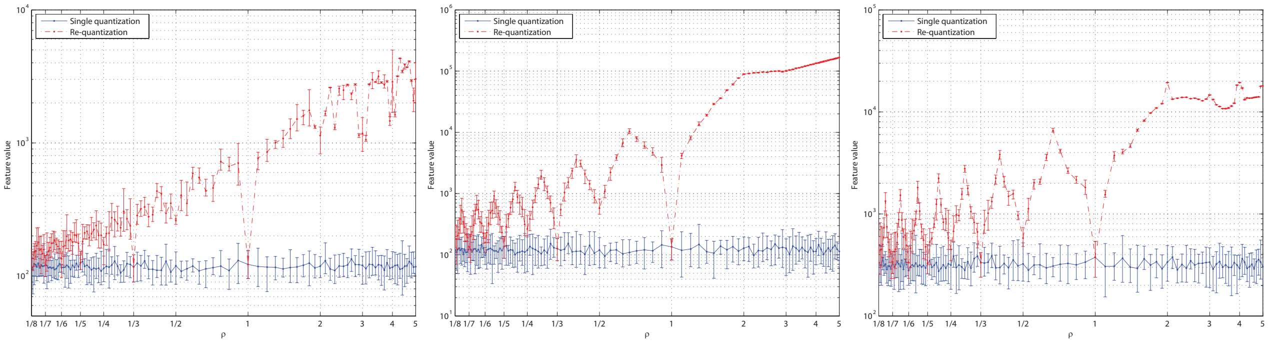
$$\mathbf{f}_2 = \mathrm{var}(\mathbf{d}(k)) \tag{7}$$

where var(.) is the variance operator. Again, in the presence of a re-quantized signal, this feature is expected to be larger than for a single quantized signal.

As already mentioned in previous works, artifacts introduced by re-quantization are periodic. This periodicity is also present in the approximation error when the input signal is re-quantized. Therefore, we introduce a last feature $\mathbf{f}_3$ which considers the maximum magnitude of the spectrum $\mathbf{D}(u)$ of the approximation error, i.e.:

$$\mathbf{f}_3 = \max_u |\mathbf{D}(u)|. \tag{8}$$

If the tested signal is re-quantized, the histogram approximation reveals some periodic patterns which translate into a strong peak in the frequency domain. On the other hand, single quantized signals do not have such periodic features and it is therefore expected that the computed feature value will be, on average, smaller.



(a) $\mathbf{f}_1$

(b) $\mathbf{f}_2$

(c) $\mathbf{f}_3$

Figure 2. Behaviour of the three proposed features for single quantized and re-quantized synthetic signals at different re-quantization ratio values. The error bars indicate the minimum and maximum values observed after 10 simulations.

3.2 Evaluation of the Features

Figure 2 depicts the behaviour of the three proposed features for several re-quantization ratios ρ. The input signal is a 10,000-samples long signal, whose samples follow a Gaussian distribution with zero mean and variance 100. The features are computed when the input signal has been subjected (i) either to a single quantization with step size $\Delta = 1$, to (ii) to re-quantization with $\Delta_2 = 1$ and ρ varying between 1/8 and 5. The error bars indicate the minimum and maximum values observed after 10 simulations. Based on the different Figures, one can see that the proposed features generally take very different values depending on whether the input signal was single quantized or re-quantized. The only exception is when the re-quantization ratio is in the neighborhood of $1/\rho \in \mathbb{N}^*$, in which case the features values are nearly the same for both type of input. Finally, the discriminative power of the three proposed features appears to vanish as the re-quantization ratio ρ decreases. All these observations are in accordance with the theoretical analysis conducted Section 2. Still, these features seem to offer the potential to construct classifiers to detect re-quantization.

Any machine learning technique can be used to build a binary classifier which would take as input the feature vector $\mathbf{f} = [\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3]$ and assess whether the tested signal is re-quantized or not. Fisher Linear Discriminant (FLD) technique described in Appendix A and Support Vector Machine (SVM) technique described in Appendix B are used to assess the ability of discriminating single quantized signals and double quantized signals by using the proposed 3-dimensional feature vector. During training, the FLD algorithm identifies a projection vector which jointly maximizes the distance between the two classes and minimizes the variance within each class. In contrast, The SVM algorithm identifies a hyperplane that can optimally separate signals into different classes. Then, during testing, both of the two algorithms give each input feature vector a value for discriminating. The values of all the testing feature vectors are then thresholded in order to decide whether the input signals are re-quantized or not. Depending on the value of the threshold τ being used, the classifier will achieve different false positive (detecting that the input signal is re-quantized when it is not) and false negative (failing to detect re-quantized signals) probabilities. By varying this threshold, one can plot a Receiver Operating Characteristic (ROC) curve which depicts the true positive probability versus the false positive probability, and subsequently extract some performance metric such as the Area Under the Curve (AUC).[14] A classifier that is equivalent to random guessing will result in an AUC value close to 0.5 whereas an accurate classifier will have an AUC value close to 1.

Figure 3 reports the classification performances achieved when applying FLD and SVM classifiers to the three features introduced in Subsection 3.1. Training has been done with 2,000 signals and testing with 8,000 signals. Simulations were repeated 10 times for cross validation and the error bars indicate the minimum and maximum AUC values observed during simulations. In our strategy, a single classifier (FLD/SVM) is trained with all possible re-quantization ratios in order to better simulate practical scenarios, i.e. the classifier has to

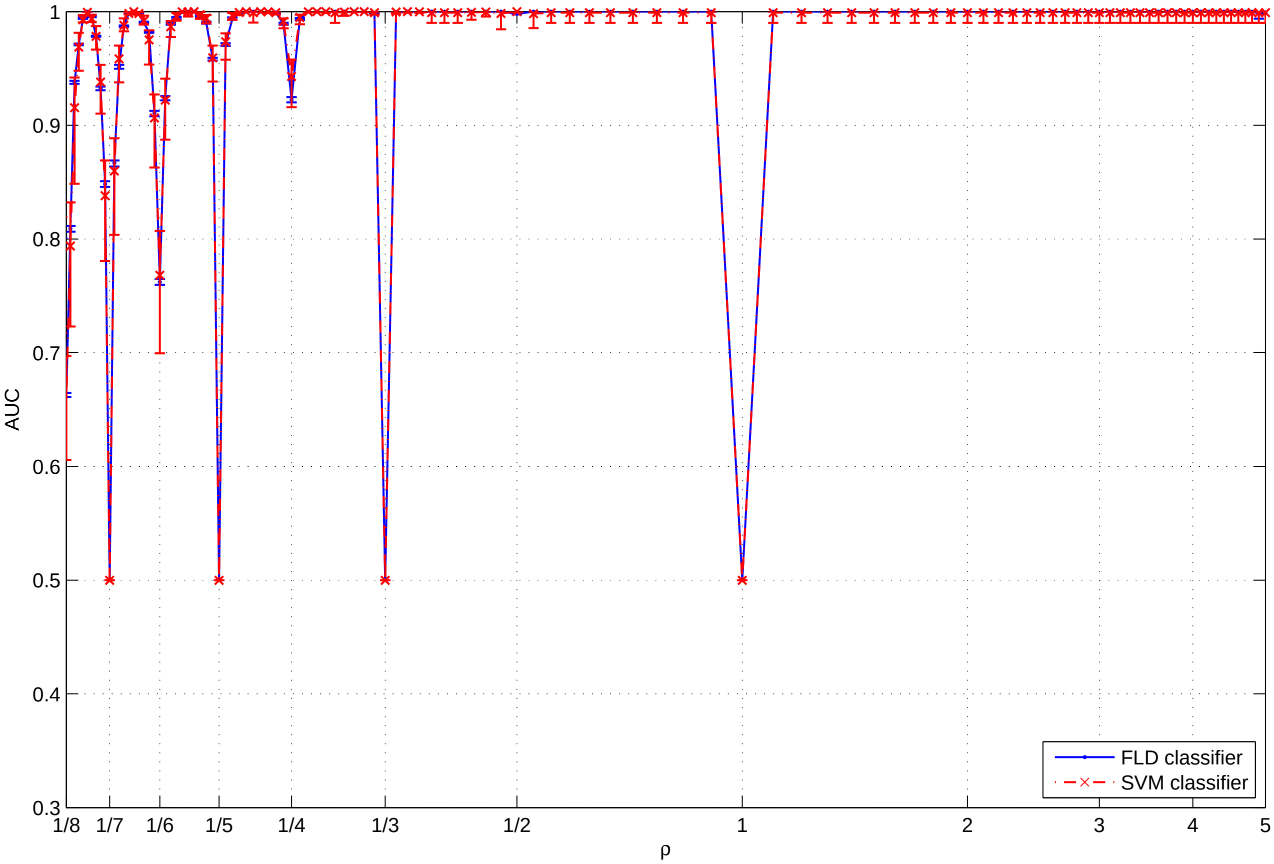


Figure 3. Performances of FLD-based and SVM-based classification on synthetic signal at different re-quantization ratios ρ. The error bars indicate the minimum and maximum values observed during simulations.

make a decision without knowing the re-quantization ratio, if there has been any re-quantization.

Figure 3 clearly suggests that it is feasible to train a single classifier for all possible re-quantization ratios. The performances of the FLD classifier and the SVM classifier are comparable with each other. However, the error bar for the FLD classifier is a bit tighter than that of SVM, which suggests that the performance of the FLD classifier might be more stable. Our second observation is that re-quantization detection is possible for the cases where $1/\rho \notin \mathbb{N}^*$. However, for the cases where $1/\rho \in \mathbb{N}^*$, re-quantization detection accuracy declines sharply as anticipated, with the somehow noticeable exception of $1/\rho \in 2\mathbb{N}^*$. After investigation, it is due to the specific quantization process that we have adopted in Equation (2c) which introduces a specific artifact in the zero-valued bin in such cases, hence allowing for accurate discrimination. For instance, consider the situation when $\rho = 1/2$. In that case, the bin corresponding to value 0 will only receive one bin from the first quantizer whereas all other bins will receive two. It is this asymmetry which allow detection of re-quantization in these cases.

4. JPEG RE-COMPRESSION DETECTION

In this Section, the performances of the proposed classification system are assessed in a more practical scenario, namely JPEG re-compression detection for still images. After briefly reviewing the JPEG compression process with a special emphasis on the quantization process, the proposed detection process is fully described as well as two baseline systems used for comparison purposes. The classification performances for JPEG re-compression detection are then reported and thoroughly analyzed.

4.1 JPEG Compression

This paper is not intended to provide an extensive overview of JPEG compression and the interested reader is redirected to[15] for further details. In summary, JPEG compression consists of the following steps:

1. the image is first converted to the YUV colour space if necessary;
2. each colour band is submitted to a 8×8 block DCT transform;
3. each DCT coefficient $\mathbf{F}(u, v)$ is quantized with a uniform quantizer whose step size $\Delta_Q(u, v)$ depends on the JPEG quality factor $Q \in \{1, 2, 3, \ldots, 100\}$ and the frequency band of the coefficient as detailed below;
4. the resulting quantization index are then encoded:
    - DC coefficients are encoded with predictive lossless coding e.g. differential pulse-code modulation (DPCM),
    - for each block, AC coefficients are scanned in zig-zag order, run-length encoded and finally entropy coded e.g. Huffman or arithmetic coding.

Obviously, in the context of this work on re-quantization detection, the third step is particularly relevant. The quantization step size is set according to the JPEG quality factor as follows:

$$\Delta_Q(u, v) = \begin{cases} \max\left(\left\lfloor \frac{100-Q}{50}\mathbf{Q}(u, v) + 0.5\right\rfloor, 1\right) & , 50 \le Q \le 100 \\ \left\lfloor \frac{50}{Q}\mathbf{Q}(u, v) + 0.5\right\rfloor & , 0 < Q < 50, \end{cases} \tag{9}$$

where $\mathbf{Q}(u, v)$ is a reference quantization table defined in the JPEG standard which accounts for the Human Visual System (HVS) i.e. high frequency coefficients are quantized with larger step size since the HVS is less sensitive to distortions in high frequencies. A typical quantization table for the luminance channel looks like:

$$\mathbf{Q} = \begin{bmatrix} 16 & 11 & 10 & 16 & 24 & 40 & 51 & 61 \\ 12 & 12 & 14 & 19 & 26 & 58 & 60 & 55 \\ 14 & 13 & 16 & 24 & 40 & 57 & 69 & 56 \\ 14 & 17 & 22 & 29 & 51 & 87 & 80 & 62 \\ 18 & 22 & 37 & 56 & 68 & 109 & 103 & 77 \\ 24 & 35 & 55 & 64 & 81 & 104 & 113 & 92 \\ 49 & 64 & 78 & 87 & 103 & 121 & 120 & 101 \\ 72 & 92 & 95 & 98 & 112 & 100 & 103 & 99 \end{bmatrix}. \tag{10}$$

4.2 Re-compression Detection

Due to the rounding operation in Equation (9), in case of re-compression, coefficients from different frequency bands may experience re-quantization with different re-quantization ratios ρ. As a result, it is advantageous to consider different frequency bands and combine them together. In other words, for an incoming image, we first compute the three features described in Section 3 for each frequency band and then combine them together, i.e. feature extraction, concatenation, and FLD-based (SVM-based) classification.

At this point, our proposed detection method processes 16 bands together. The 16 bands are selected as being the 4×4 upper left corner bands. The tree proposed features $\mathbf{f}_{i,j} = [\mathbf{f}_1(i, j),\ \mathbf{f}_2(i, j),\ \mathbf{f}_3(i, j)]$ are computed for each one of these 16 bands and are concatenated as a $3 \times 16 = 48$ dimensional feature vector $\mathbf{F}$, i.e.

$$\mathbf{F} = \Big\{\mathbf{f}_{i,j} | (i, j) \in \mathcal{B}\Big\}. \tag{11}$$

where $\mathcal{B}$ represents the selected frequency bands, indexed by i and j according to their rows and columns in the block DCT decomposition. Each image is thus reduced to a 48-D feature vector which can be subsequently used for training and/or testing classifiers.

For comparison purposes, two state-of-the-art baseline algorithms are also considered.

4.2.1 Baseline 1: Histogram features[12]

For the selected frequency bands, the 16 first buckets of the signal histogram are concatenated together, i.e.

$$\mathbf{f}_{i,j}^{\mathrm{B1}} = \Big\{ \frac{1}{C_{i,j}} (\mathbf{h}_{i,j}(0), \ldots, \mathbf{h}_{i,j}(15)) | (i,j) \in \mathcal{B} \Big\}, \tag{12}$$

where $\mathbf{h}_{i,j}(x)(x = 0, \ldots, 15)$ represents the x^{th} bucket of the histogram for the frequency bands (i, j), and $C_{i,j}$ is the normalization factor defined as follows:

$$C_{i,j} = \sum_{x=0}^{15} \mathbf{h}_{i,j}(x). \tag{13}$$

The features obtained for the selected set of frequency bands $\mathcal{B}$ are then concatenated in a single feature vector:

$$\mathbf{F}_{\mathrm{B1}} = \Big\{ (\mathbf{f}_{i,j}^{\mathrm{B1}} | (i,j) \in \mathcal{B} \Big\}. \tag{14}$$

resulting in a single 256-D feature vector.

4.2.2 Baseline 2: Generalized Benford's Law features[13]

The Generalized Benford's Law (GBL) models the distribution of the first digits of JPEG DCT coefficients by a parametric logarithmic function as

$$\mathrm{P}(x) = N \times \log \left(1 + \frac{1}{s + x^q} \right), \; x = 1, \ldots, 9. \tag{15}$$

where $\mathrm{P}(x)$ stands for the probability that the first digit of a JPEG DCT coefficient is equal to x; N is a normalization factor which makes $\mathrm{P}(x)$ a probability distribution; s and q are the model parameters which may vary for different images and different compression Q-factors.

Given the histogram of a particular frequency band of a JPEG image, it is possible to estimate the parameters $\hat{s}$ and $\hat{q}$ so that the model given in Equation (15) best fits the provided data. For that particular frequency band (i, j), one can then look at how much the observed data deviates from the estimated model:

$$\mathbf{f}_{i,j}^{\mathrm{B2}} = \sum_{x=1}^{9} (\hat{\mathrm{P}}_{i,j}(x) - \mathrm{P}_{i,j}(x))^2. \tag{16}$$

Relying on the assumption that a single-compressed image should follow the model pretty well whereas a double compressed one is likely to deviate from it, this feature provides input material for classification.

For the selected frequency bands $\mathcal{B}$, the Generalized Benford's Law feature vector is obtained by concatenating the individual features for each frequency band:

$$\mathbf{F}_{\mathrm{B2}} = \Big\{ \mathbf{f}_{i,j}^{\mathrm{B2}} | (i,j) \in \mathcal{B} \Big\}. \tag{17}$$

resulting in a 16-D feature vector.

4.3 Experimental Evaluation

In our experiments, we used a database of 10,000 never-compressed 512 × 512 images. In case of single compression, the quality factor $Q = 75$ has been used and, for double compression, we set the quality factor of the second compression to $Q_2 = 75$. To evaluate the classifications performances, 20% of the images were used for training and the remainder 80% for testing. All simulations are repeated 10 times for cross-validation.

Figure 4 compares the performances of our detection system with the two baseline systems using the FLD-based classification. As could be anticipated, JPEG re-compression is easily detected when $Q_1 < Q_2$. However,

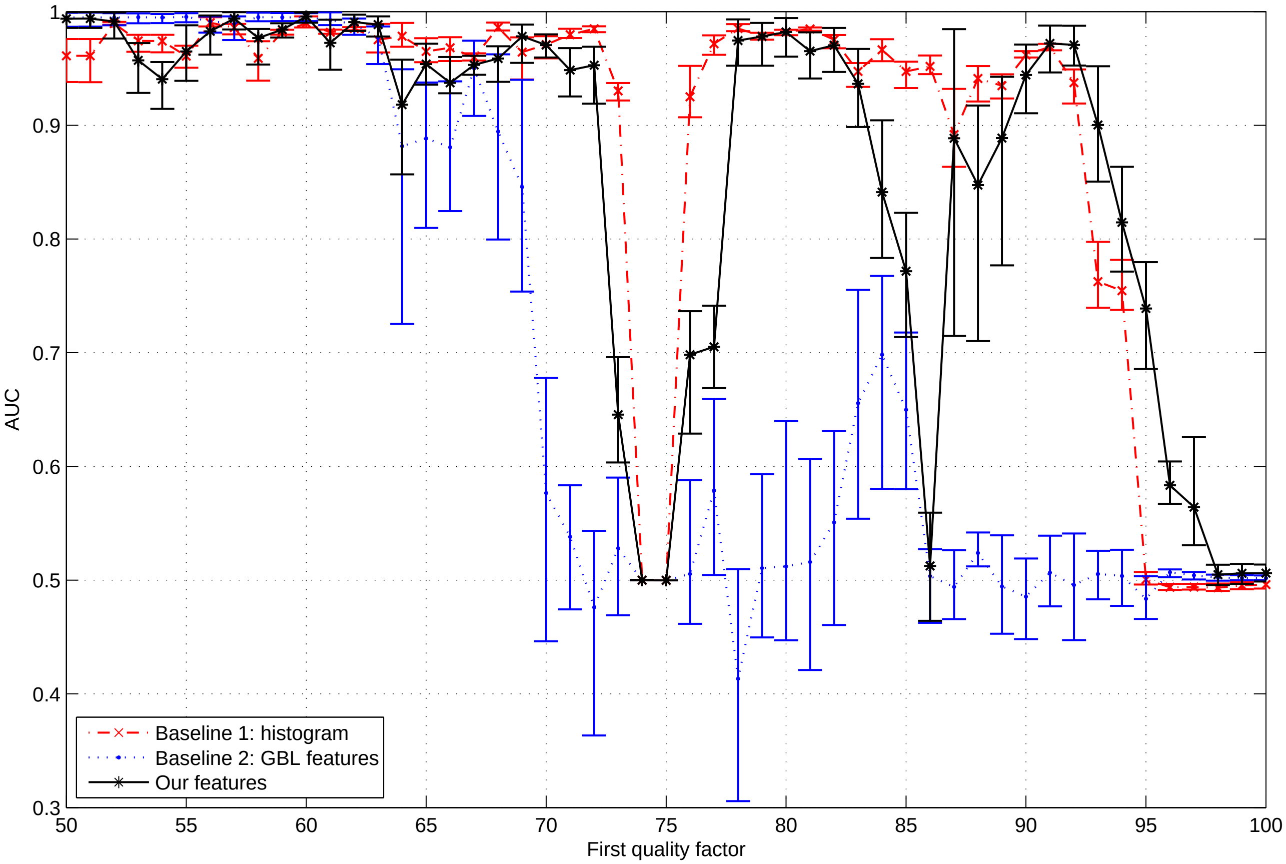


Figure 4. Performances of JPEG re-compression detection using the FLD-based classification system. The error bars indicate the observed minimum and maximum values during simulations.

the proposed system fails to deal with specific re-compression ratios ($Q_1 = 74, 97, 98, \ldots$) which basically correspond to situations where the re-quantization ratios for most of the considered coefficients have the form $1/\rho \in \mathbb{N}^*$. Moreover, with FLD-based classification, the error bars appear to be quite loose, thus reflecting the limitation of FLD classification in terms of stability in practice. Overall, the proposed system outperforms the GBL-based algorithm and exhibits somehow similar classification performances compared to the histogram features system. The main differences between these two systems can be summarized as follows: the histogram features system outclasses the proposed features by a large margin around $Q = 85$ whereas the proposed system is slightly better for high quality factors Q. More unexpectedly, significant performances degradation are observed in general when experiments are conducted on real images, compared with experiments conducted on synthetic data reported in Figure 3. This is likely to be due to the fact that JPEG DCT coefficients do not follow a Gaussian distribution. This phenomenon requires further research and analysis.

The experiments have then been repeated using SVM classification in Figure 5 and again performances are compared to those of the two baseline systems. In contrast with the observations for synthetic experiments, the classification technique now clearly has a strong impact on performances. Overall, classification results with SVM are much better, although the second baseline system is still far from acceptable performances. The proposed feature set now has very similar classification performances compared to the baseline system which uses the histogram features. In particular, the weakness around $Q = 85$ seems no longer to hold.

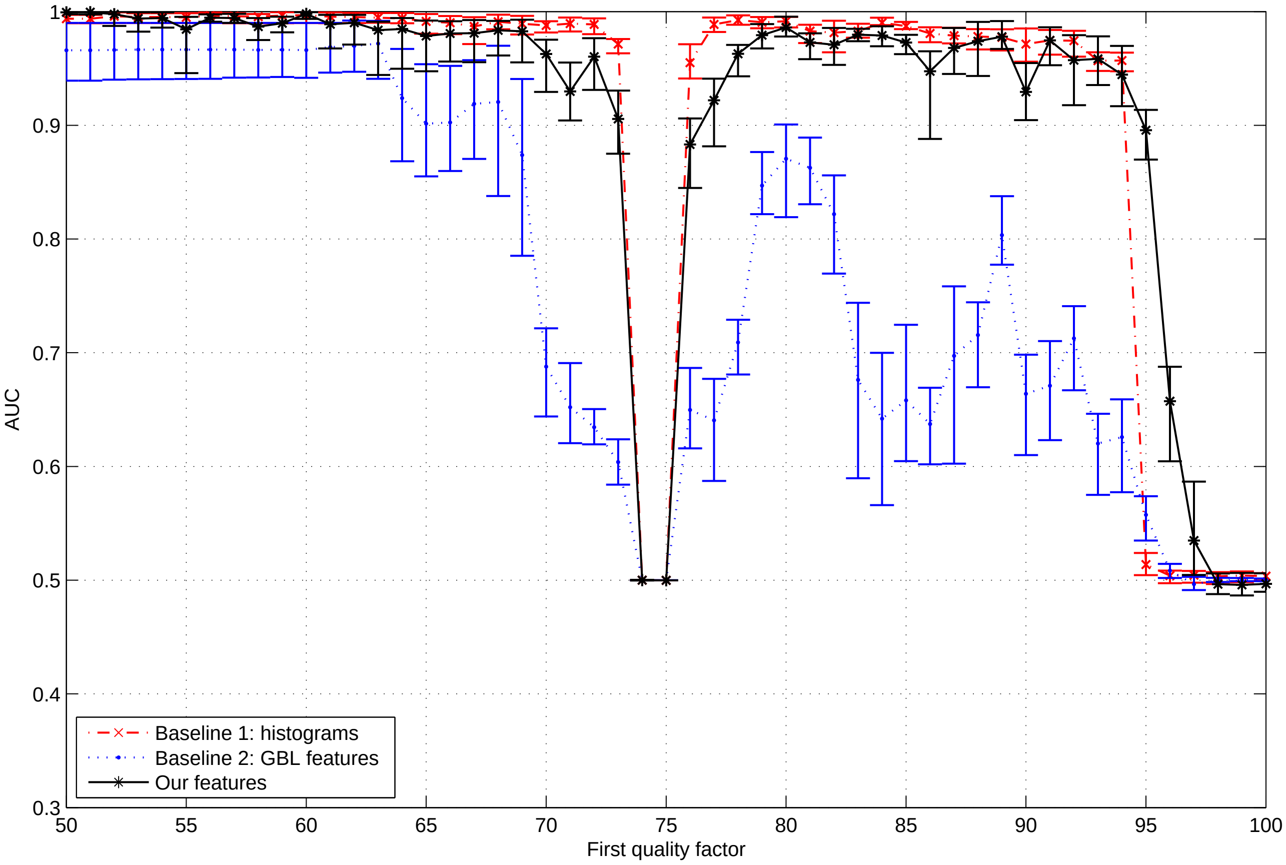


Figure 5. Performances of JPEG re-compression detection using the SVM-based classification system. The error bars indicate the observed minimum and maximum values during simulations.

5. CONCLUSIONS AND FUTURE WORKS

In this paper, we first introduced 3 features to detect re-quantization. These features were derived from the fact that the re-quantization process introduces (i) periodic artifacts and (ii) discontinuities in the histogram. The discriminative power of these features has then been verified on synthetic data.

Based on these encouraging classification results, we then focused on detecting JPEG re-compression for still images and exploited two baseline systems for comparison purposes. Both FLD and SVM classification have been investigated. The reported experimental results clearly demonstrated of the proposed system to match the classification performances of ealier algorithms but with a significantly reduced feature set. As with the other systems, re-quantization detection still fails in very specific cases where the re-compression process does not introduces statistical artifacts.

Future work will investigate alternative classification strategies e.g. detecting re-quantization based on anomaly detection, and explore the ability of such systems to deal with multiple quantizations, e.g. in the context of complex multimedia system such as voice transmission over the telephony network.

APPENDIX A. FISHER LINEAR DISCRIMINANT ANALYSIS

Fisher Linear Discriminant (FLD) analysis identifies vector orientations that are efficient for discrimination. The goal is to find an orientation $\mathbf{u}$ for which the samples in the dataset, once projected onto it, are well separated.

Let us assume that a dataset $\mathcal{D}$ is made of N d-dimensional samples $\mathbf{x_1}, \ldots, \mathbf{x}_N$, N_1 being in a subset $\mathcal{D}_1$ corresponding to one class and N_2 being in a subset $\mathcal{D}_2$ corresponding to the other class. The first step of FLD analysis consists in computing the d-dimensional sample mean of each class:

$$\mathbf{m}_i = \frac{1}{N_i} \sum_{\mathbf{x} \in \mathcal{D}_i} \mathbf{x}. \tag{18}$$

Next, the scatter matrix $\mathbf{S}_W = \mathbf{S}_1 + \mathbf{S}_2$ is computed using the following definitions:

$$\mathbf{S}_i = \sum_{\mathbf{x} \in \mathcal{D}_i} (\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^t. \tag{19}$$

Finally, the direction of projection $\mathbf{u}$ is given by:

$$\mathbf{u} = \mathbf{S}_W^{-1}(\mathbf{m}_1 - \mathbf{m}_2) = \mathbf{S}_W^{-1}\mathbf{S}_B. \tag{20}$$

This vector $\mathbf{u}$ defines a linear function $y = \mathbf{u}^t\mathbf{x}$ which yields the maximum ratio of between-class scatter ($\mathbf{S}_B$) to within-class scatter ($\mathbf{S}_W$). The interested reader is redirected to[16] for further details (pp. 117–121).

APPENDIX B. SUPPORT VECTOR MACHINE

For d-dimensional samples, Support Vector Machine (SVM) firstly maps the input d-dimensional samples to a higher k-dimensional space ($k \geq d$) by applying different kernel functions, and then identifies a $(k-1)$-dimensional hyperplane that can optimally separate the transformed samples into different classes.

The separating hyperplane of SVM is decided such that the larger the margin between the two edge lines is, the better is the separating line. The input vectors are labeled as a pair $(\mathbf{x}_i, \mathbf{y}_i)$, $i = 1, 2, \cdot \cdot \cdot, n$, $\mathbf{x}_i \in \mathbf{R}^d$, $\mathbf{y}_i \in \{-1, 1\}$. The vectors lying on the hyperplane satisfy the function $\mathbf{w} \cdot \mathbf{x} + b = 0$, where $\mathbf{w}$ is normal to the hyperplane. All the other vectors satisfy the constraints below:

$$\mathbf{x}_i \cdot \mathbf{w} + b \geqq \mathbf{y}_i, \quad \text{when } \mathbf{y}_i = 1 \tag{21}$$

$$\mathbf{x}_i \cdot \mathbf{w} + b \leqq \mathbf{y}_i, \quad \text{when } \mathbf{y}_i = -1 \tag{22}$$

When $\mathbf{x}_i$ falls on the edge lines, i.e. $\mathbf{H}_1$, $\mathbf{H}_2$, the formulas above reduce to equalities. Furthermore, these two formulas can be combined into one as follows:

$$\forall i,\ \mathbf{y}_i \cdot (\mathbf{x}_i \cdot \mathbf{w} + b) - 1 \geqq 0 \tag{23}$$

It is obvious that the perpendicular distance from origin point to $\mathbf{H}_1$ is $\frac{|1-b|}{||\mathbf{w}||}$, while that of $\mathbf{H}_2$ is $\frac{|-1-b|}{||\mathbf{w}||}$. Thus, the perpendicular distance between $\mathbf{H}_1$ and $\mathbf{H}_2$ is $\frac{2}{||\mathbf{w}||}$. So the problem of finding a SVM hyperplane with maximum margin between $\mathbf{H}_1$ and $\mathbf{H}_2$ becomes minimizing $||\mathbf{w}||^2$. SVM then switches to a Lagrangian formulation to solve the optimization problem as

$$\mathrm{L}(\mathbf{w}, \mathbf{a}) = \frac{1}{2}||\mathbf{w}||^2 - \sum_{i=1}^{n} \mathbf{a}_i\mathbf{y}_i(\mathbf{x}_i \cdot \mathbf{w} + b) + \sum_{i=1}^{n} \mathbf{a}_i \tag{24}$$

Given the objective function as a quadratic function, according to Lagrangian formulation, there is only one solution to the above function. Kuhn and Tucker pointed out that the necessary condition is:

$$\frac{\partial \mathrm{L}(\mathbf{w}, \mathbf{a})}{\partial \mathbf{w}} = \mathbf{w} - \sum_{i=1}^{n} \mathbf{a}_i\mathbf{y}_i\mathbf{x}_i = 0 \tag{25}$$

$$\frac{\partial \mathrm{L}(\mathbf{w}, \mathbf{a})}{\partial b} = -\sum_{i=1}^{n} \mathbf{a}_i\mathbf{y}_i = 0 \tag{26}$$

$$\mathbf{y}_i(\mathbf{x}_i \cdot \mathbf{w} + b) - 1 \geq 0 \tag{27}$$

$$\mathbf{a}_i(\mathbf{y}_i(\mathbf{x}_i \cdot \mathbf{w} + b) - 1) \geq 0 \tag{28}$$

$$\mathbf{a}_i \geq 0 \tag{29}$$

Above Support Vector Machine defines a simple separable case. For more complicated cases, interested readers are redirected to[16] for further details.

ACKNOWLEDGMENTS

Ms. Xiaoying Feng's research is jointly funded by the British Engineering and Physical Sciences Research Council (EPSRC) and British Telecommunications plc. (BT) through a Dorothy Hodgkin Postgraduate Award.

REFERENCES

[1] H. Farid, "Photo tampering throughout history."
[online] http://www.cs.dartmouth.edu/farid/research/digitaltampering.

[2] R. Descartes, *Metaphysical Meditations*, 1641.

[3] G. Palmer, "A road map for digital forensic research," Tech. Rep. T001-01, Digital Forensic Research Workshop, August 2001.

[4] H. Farid, "Image forgery detection – A survey," *IEEE Signal Processing Magazine* **26**, pp. 16–25, March 2009.

[5] A. Westfeld, "High capacity despite better steganalysis (F5 a steganographic algorithm)," in *Proceedings of the 4th Information Hiding Workshop*, I. S. Moskowitz, ed., *Lecture Notes in Computer Science* **2137**, pp. 289–302, April 2001.

[6] N. Provos, "Defending against statistical steganalysis," in *Proceedings of the 10th USENIX Security Symposium*, August 2001.

[7] J. Fridrich, M. Goljan, and D. Hogea, "Steganalysis of JPEG images: Breaking the F5 algorithm," in *Proceedings of the 5th Information Hiding Workshop*, F. A. P. Petitcolas, ed., *Lecture Notes in Computer Science* **3200**, pp. 310–323, October 2002.

[8] J. He, Z. Lin, L. Wang, and X. Tang, "Detecting doctored JPEG images via DCT coefficient analysis," in *Proceedings of the 9th European Conference on Computer Vision, Part III*, A. Leonardis, H. Bischof, and A. Pinz, eds., *Lecture Notes in Computer Science* **3953**, pp. 423–435, May 2006.

[9] A. C. Popescu and H. Farid, "Statistical tools for digital forensics," in *Proceedings of the 6th Information Hiding Workshop*, J. Fridrich, ed., *Lecture Notes in Computer Science* **3200**, pp. 128–147, May 2004.

[10] W. Wang and H. Farid, "Exposing digital forgeries in video by detecting double MPEG compression," in *Proceedings of the 8th ACM Workshop on Multimedia and Security*, pp. 37–47, September 2006.

[11] J. Lukáš and J. Fridrich, "Estimation of primary quantization matrix in double compressed JPEG images," in *Proceedings of the Digital Forensic Research Workshop*, August 2003.

[12] T. Pevný and J. Fridrich, "Detection of double-compression for applications in steganography," *IEEE Transactions on Information Forensics and Security* **3**, pp. 247–258, June 2008.

[13] D. Fu, Y. Q. Shi, and W. Su, "A generalized Benford's law for JPEG coefficients and its applications in image forensics," in *Security and Watermarking of Multimedia Contents IX*, *Proceedings of SPIE* **6505**, pp. 1L1–1L11, January 2007.

[14] T. Fawcett, "ROC graphs: Notes and practical considerations for researchers," tech. rep., HP Laboratories, March 2004.

[15] G. K. Wallace, "The JPEG still picture compression standard," *IEEE Transactions on Consumer Electronics* **38**, pp. 18–34, February 1992.

[16] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, Wiley-Interscience, 2nd ed., 2001.