

AN ENERGY-BASED METHOD FOR THE FORENSIC DETECTION OF RE-SAMPLED IMAGES

Xiaoying Feng¹, Ingemar J. Cox¹, and Gwenaël Doërr²

¹ Department of Computer Science, University College London, UK
{x.feng, i.cox}@cs.ucl.ac.uk

² Security & Content Protection Labs, Technicolor R&D France
gwenaël.doerr@technicolor.com

ABSTRACT

We propose a new method to detect re-sampled imagery. The method is based on examining the normalized energy density present within windows of varying size in the second derivative of the frequency domain, and exploiting this characteristic to derive a 19-dimensional feature vector that is used to train a SVM classifier. Experimental results are reported on 7,500 raw images from the BOSS database. Comparison with prior work reveals that the proposed algorithm performs similarly for re-sampling rates greater than 1, and is superior to prior work for re-sampling rates less than 1. Experiments are performed for both bilinear and bicubic interpolation, and qualitatively similar results are observed for each. Results are also provided for the detection of re-sampled imagery that subsequently undergoes JPEG compression. Results are quantitatively similar with some small degradation in performance as the quality factor is reduced.

Index Terms— Image forensics, Re-sampling detection, Normalized energy density.

1. INTRODUCTION

Forensic signal processing attempts to identify the variety of processing steps that a signal has undergone. Such information is useful in determining whether, for example, a signal is authentic or has been tampered with. There are two main approaches to multimedia forensics: active forensics and passive forensics [1]. Active forensics relies on modifying the multimedia signal prior to its distribution to assist in later forensic analysis. Digital watermarks [2] are one example of active forensics. A limitation of active forensics is the need for content-generating devices, e.g. cameras, sensors, microphones, to be capable of embedding watermarks. This is often not possible, and in these cases active forensics cannot be applied. Passive forensics, in contrast, does not rely on any prior modification of the signal. As such, passive forensics is, in theory, applicable to a broader range of operating scenarios.

A variety of passive forensic methods have been developed to detect, for example, re-quantization [3, 4, 5, 6], re-

sampling [7, 8, 9, 10, 11, 12], and affine transformations [13, 14]. In this paper, we consider the problem of determining whether an image* has undergone re-sampling. Section 2 provides a review of prior work on this topic. Section 3 then describes our proposed algorithm in detail. In Section 4, experiments are performed using the BOSS database [15] consisting of 7,500 raw images, and compared with the algorithm described in [9]. Finally, Section 5 summarizes our results and discusses possible directions for future work.

2. PRIOR WORK

Early work on detecting image re-sampling was based on the observation that artifacts were introduced in the re-sampled images due to interpolation, a basic operation involved in re-sampling. Generally, these artifacts are periodic in the spatial domain and therefore manifest themselves as peaks in the corresponding frequency domain.

Popescu and Farid [7] noted that the interpolation process introduces correlations between the re-sampled image pixels. They proposed measuring these correlations based on an expectation/maximization (EM) algorithm. The EM algorithm estimates the linear correlation between each pixel and its neighbors, and subsequently computes the probability of each pixel being correlated to its neighbors. The corresponding correlation probability map (p-map) in the DFT domain, exhibits periodic peaks that are not present in single-sampled images.

The work of Popescu and Farid was subsequently refined. Mahdian and Saic [8, 9] proposed an automatic detector that searches for local maxima, which is defined as n times greater than a local average magnitude. Kirchner [10] replaced the EM algorithm, which is computationally demanding, with linear filtering and cumulative periodograms, and proposed an automatic detector based on the maximum gradient of the p-

*For the sake of simplicity, we only discuss grey-scale images in this paper. However, a color image can be represented by three channels, i.e. intensity, saturation and hue. Our proposed method can be directly applied in the intensity channel of color images.

map. Besides detecting re-sampled raw camera images, Gallagher [11] and Kirchner and Gloe [12] investigated the detection of re-sampling on JPEG compressed images.

The method described in [9] is used as the baseline against which we compare our algorithm. The second derivative is taken along either the horizontal or the vertical dimension of an image. The radon transformation then projects the second derivative to each of 180 directions, where the projection angles are integers from 0° to 179° . The detection of re-sampling is based on the detection of periodicity in the autocovariance of the projected vectors. In so doing, the first derivative of all 180 projected vectors are calculated. The autocovariance of the first derivative is then computed. Finally, the periodicity of autocovariance is detected in the DFT domain, using a local maxima detector. Further details of the algorithm can be found in [9].

3. NORMALIZED ENERGY DENSITY-BASED METHOD

Our method for detecting re-sampling is based on a 19-dimensional feature vector that is derived from an examination of the normalized energy density present at various window sizes in the DFT of the second-derivative of an image.

Let $\mathbf{i}(x, y)$ denote a single-sampled or a re-sampled image. Its corresponding DFT is denoted by $\mathbf{I}(u, v)$. We assume, for simplicity and without loss of generality, that the image is a square with dimension N , i.e. there are N^2 pixels present in the image.

Using Parseval's equation, the energy, E , present in an image is given by:

$$E = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} \mathbf{i}(x, y)^2 = \sum_{u=-f_c}^{f_c} \sum_{v=-f_c}^{f_c} |\mathbf{I}(u, v)|^2, \quad (1)$$

where f_c is the cutoff frequency of the image.

We define $E(w)$ to be the energy present in the power spectrum of an image, in a window of dimension $\pm w$, where $0 < w \leq f_c$:

$$E(w) = \sum_{u=-w}^w \sum_{v=-w}^w |\mathbf{I}(u, v)|^2. \quad (2)$$

Since an original image and a resized image will not have the same dimension, we normalize the window size, s , with respect to the cutoff frequency, f_c , where s takes values between 0 and 1. Thus, the energy, $E(s)$, present in a window s is given by:

$$E(s) = \sum_{u=-s \cdot f_c}^{s \cdot f_c} \sum_{v=-s \cdot f_c}^{s \cdot f_c} |\mathbf{I}(u, v)|^2. \quad (3)$$

It is well-known that the energy present in an image is typically concentrated in the lower frequencies, and drops off

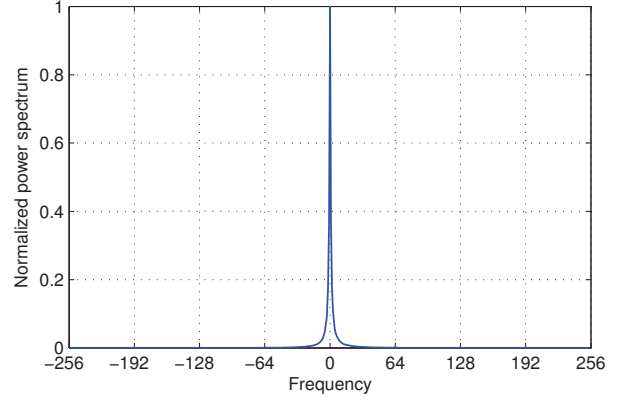


Fig. 1. The normalized power spectrum in the horizontal direction (i.e. $v = 0$). The curve is averaged over 7,500 images from the BOSS database.

with increasing frequency. The shape of the power spectrum is often approximated by a Gaussian or Laplacian model. Figure 1 shows the normalized power spectrum in the horizontal direction, i.e. $|\mathbf{I}(u, 0)|^2$, averaged over all 7,500 images in the database we used.

In practice, the high concentration of energy in the lower frequencies can mask high frequency effects. Thus, to highlight these, an image, $\mathbf{i}(x, y)$, is first high-pass filtered before examining its power spectrum. The high-pass filter with a kernel \mathbf{h} , shown in Equation 4, is used in our experiments.

$$\mathbf{h} = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}. \quad (4)$$

An image after the high-pass filter is denoted by $\tilde{\mathbf{i}}(x, y)$, and its power spectrum is denoted by $|\tilde{\mathbf{I}}(u, v)|^2$. The energy, $\tilde{E}(s)$, of an image after the high-pass filter is given by:

$$\tilde{E}(s) = \sum_{u=-s \cdot f_c}^{s \cdot f_c} \sum_{v=-s \cdot f_c}^{s \cdot f_c} |\tilde{\mathbf{I}}(u, v)|^2. \quad (5)$$

The energy density, $\tilde{E}_d(s)$, is then defined by the averaged energy within the window s , i.e.

$$\tilde{E}_d(s) = \frac{1}{(2s \cdot f_c)^2} \sum_{u=-s \cdot f_c}^{s \cdot f_c} \sum_{v=-s \cdot f_c}^{s \cdot f_c} |\tilde{\mathbf{I}}(u, v)|^2. \quad (6)$$

Finally, the normalized energy density, $\tilde{E}_n(s)$, is given by:

$$\begin{aligned} \tilde{E}_n(s) &= \frac{\tilde{E}_d(s)}{\tilde{E}_d(1)} \\ &= \frac{1}{s^2} \frac{\sum_{u=-s \cdot f_c}^{s \cdot f_c} \sum_{v=-s \cdot f_c}^{s \cdot f_c} |\tilde{\mathbf{I}}(u, v)|^2}{\sum_{u=-f_c}^{f_c} \sum_{v=-f_c}^{f_c} |\tilde{\mathbf{I}}(u, v)|^2}, \end{aligned} \quad (7)$$

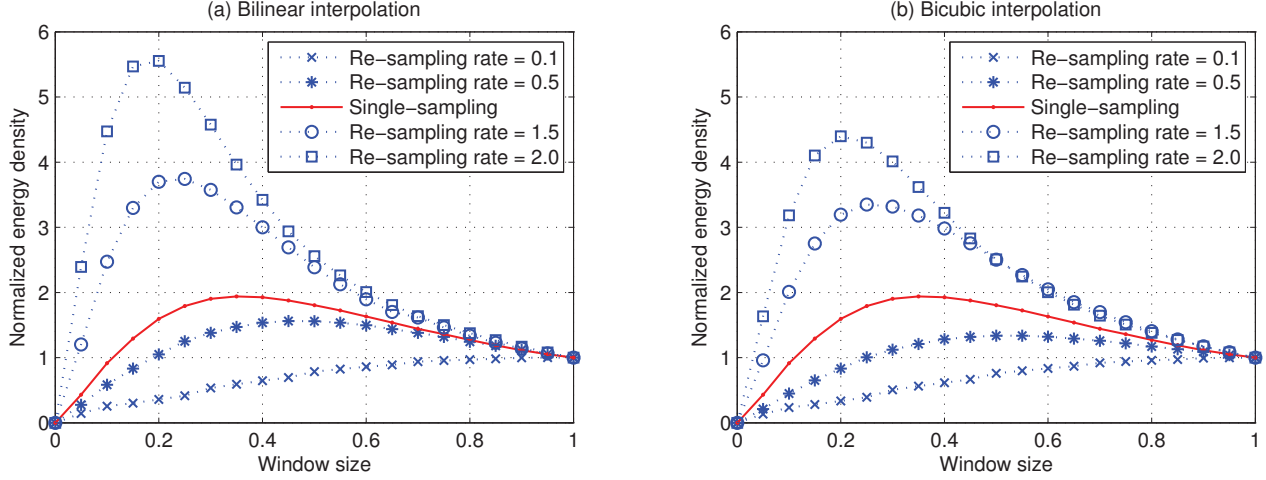


Fig. 2. The normalized energy density (without anti-aliasing effect) averaged over 7,500 images from the BOSS database, for a set of re-sampling rates.

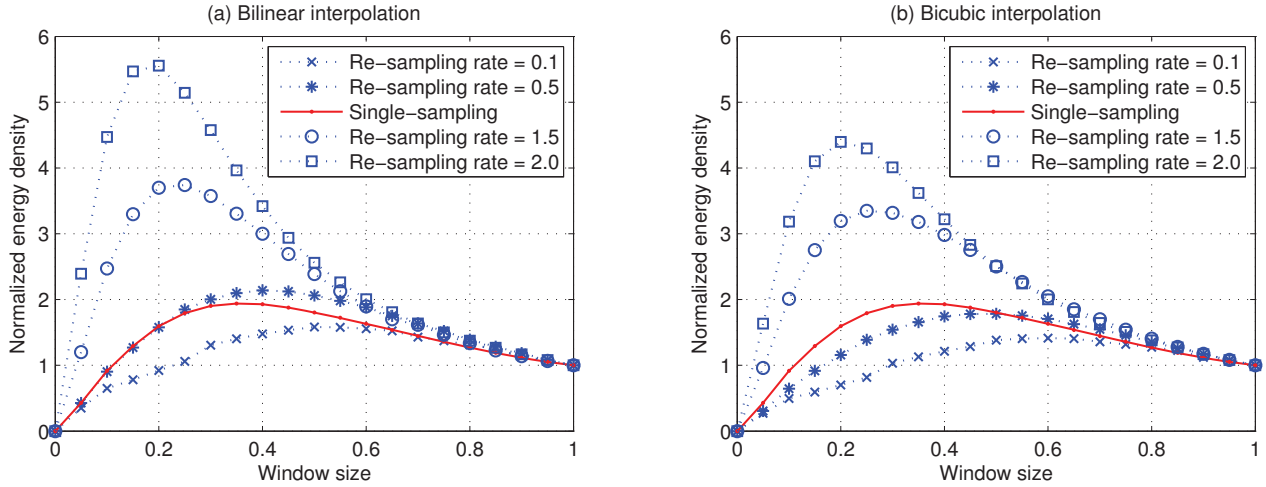


Fig. 3. The normalized energy density (with anti-aliasing effect) averaged over 7,500 images from the BOSS database, for a set of re-sampling rates.

where $\tilde{E}_d(1)$ stands for the energy density of the whole image.

Figure 2 shows the normalized energy density over 7,500 images from the BOSS database for window sizes ranging from 0 to 1, including (i) single-sampled images (solid curve); (ii) re-sampled images up-sampled at rates 1.5 and 2.0; and (iii) re-sampled images down-sampled at rates 0.1 and 0.5.

We observe the following. First, for the single-sampled imagery, we see a curve originating at zero (since there is no energy in a window size of zero), increasing monotonically to a peak at a window size of approximately $s = 0.4$, and then monotonically decreasing to a value of 1 when the window size encompasses the entire power spectrum ($s = 1$). If we now compare this to the up-sampled images, we observe a qualitatively similar curve, but the peak values are shifted to the left, i.e. the peak values occur for $s < 0.4$. Similarly for

the down-sampled images, we observe a qualitatively similar curve, but the peak values are shifted to the right, i.e. the peak values occur for $s > 0.4$. This is to be expected since, for up-sampled imagery, we would expect that less energy is present in the very high frequencies, as these frequencies are absent in the original single-sampled images. Conversely, for down-sampled imagery, the fact that we have more energy in the higher frequencies indicates that these images were derived from images containing higher frequency information.

In practice, anti-aliasing is routinely applied after the interpolation operation. Figure 3 shows the equivalent energy curves to Figure 2, after anti-aliasing has been applied. Once again, the normalized energy density curves (with anti-aliasing effect) of different re-sampling rates are well separated from each other.

The curves in both Figures 2 and 3 suggest that it is pos-

sible to differentiate single-sampled and re-sampled imagery based on these characteristics. To investigate this, we defined a 19-dimensional vector, the values of each dimension being the normalized energy density, $\tilde{E}_n(s)$, for window size s ranging from 0.05 to 0.95 in steps of 0.05.

4. EXPERIMENTS

Our experiments use the BOSS database [15], which consists of 7,500 raw images, i.e. the images have never undergone re-sampling. Experiments are performed for both bilinear and bicubic interpolation. Note that both interpolation and anti-aliasing are applied to the images, as this is a more realistic operating scenario. Further note that all re-sampling operations involve re-sizing the images, i.e. we do not consider image rotation.

The 19-dimensional vector was used as input to train support vector machine (SVM) classifiers [16]. We applied random sub-sampling validation, where for each of the 10 trials, training is performed on a random 20% subset of the database, and testing on the remaining 80%. The reported results are the average of the 10 trials.

Instead of manually setting a fixed threshold, a receiver operating characteristics (ROC) curve [17], the threshold of which is varied, is used to evaluate our detection results. Moreover, in order to reduce performance to a single scalar number, we report the area under the ROC curve (AUC). Note that an AUC value of 0 means the detection is always false, whereas an AUC value of 1 means a perfect detection. An AUC value of 0.5 represents random guessing, which is reflected by the diagonal line between (0, 0) and (1, 1) in the ROC curve.

The method described in [9] is used as the baseline algorithm for comparative purposes. Baseline # 1 is fully implemented according to [9]. In baseline # 2, the discriminating features are implemented according to [9], but detection is based on the SVM classifier that is the same as our method. The purpose of baseline # 2 is to help identify if the performance variation is due to different discriminating features or different detectors.

We first examine the case where one classifier is trained for each specific re-sampling rate. We then consider the case of only a single classifier trained to detect all re-sampling rates. Finally, the detection of re-sampled images that have subsequently been JPEG compressed is investigated.

4.1. One Classifier per Re-sampling Rate

In the first experiment, one SVM-based classifier is trained for each individual re-sampling rate, which ranges from 0.1 to 2 in steps of 0.1. As a result, there is a total of 19 classifiers in all. The experimental results are shown in Figure 4 for both bilinear and bicubic interpolation. Note that during testing,

the re-sampled image is only presented to its corresponding classifier.

In general, the detection performance of our method is comparable to that of the baseline algorithm described in [9] when re-sampling rates are greater than 1 (i.e. the re-sampled image is larger than the single-sampled one). Note that for this case, our method never reaches perfect detection, whereas the baseline # 2 does. In contrast, when re-sampling rates are less than 1, our method performs better than the baseline algorithms. Note that the baseline # 1 actually performs worse than random, i.e. AUC values are less than 0.5. This implies that the baseline # 1 can discriminate between single-sampling and re-sampling, but the decision is opposite to the truth. This characteristic was also observed in [18], where it was suggested that a two-sided hypothesis test is actually needed. For baseline # 2, there is no such problem. Still, for down-sampling, the normalized energy density approach is almost always superior. In addition, we note that the error bars for all re-sampling rates are small, indicating that the proposed method is stable across all re-sampling rates. There is little difference in performance when either bilinear or bicubic interpolation is used.

4.2. A Single Classifier for All Re-sampling Rates

We now consider a single SVM-based classifier for all re-sampling rates. This is a more realistic scenario where there is no prior knowledge of the re-sampling rate.

If an image is used for training (i.e. within the 20% subset of the database), its single-sampled version and only one of its re-sampled versions are fed into the SVM-based classifier. This guarantees that the numbers of single-sampled and re-sampled images used to train the SVM-based classifier are the same. In addition, during training, the same number of re-sampled images for each of the 19 re-sampling rates is guaranteed.

As expected, the absolute detection performance of both our method and the baseline algorithm # 2 declines slightly, as shown in Figure 5. However, the degradation in performance is small. It is also noted that the relative performance of the two algorithms remains qualitatively the same. Once again, the error bars for all re-sampling rates are small, demonstrating the stability of the proposed algorithm. Note that (i) we do not show baseline # 1 as this curve is identical to that in Figure 4 since there is no training and (ii) the scale of vertical axis in Figure 5 is different from that in Figure 4.

4.3. Detection after JPEG Compression

In practice, an image may be compressed prior to testing. In order to address this scenario, JPEG compression with different quality factors was applied. Specifically, we consider four cases, i.e. (i) JPEG quality factor of 85; (ii) JPEG quality factor of 55; (iii) JPEG quality factor of 25; and (iv) JPEG quality factor of 10.

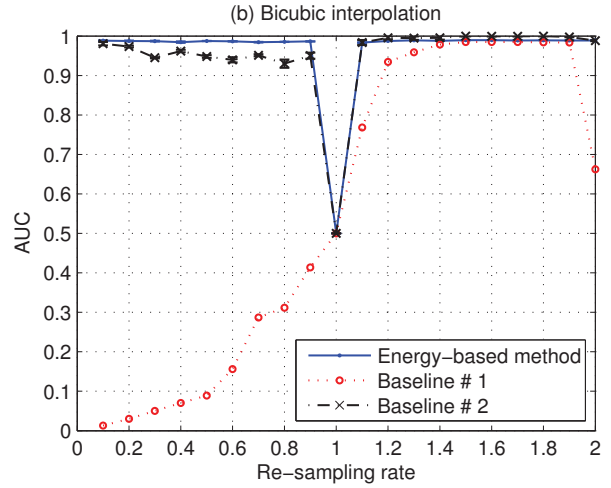
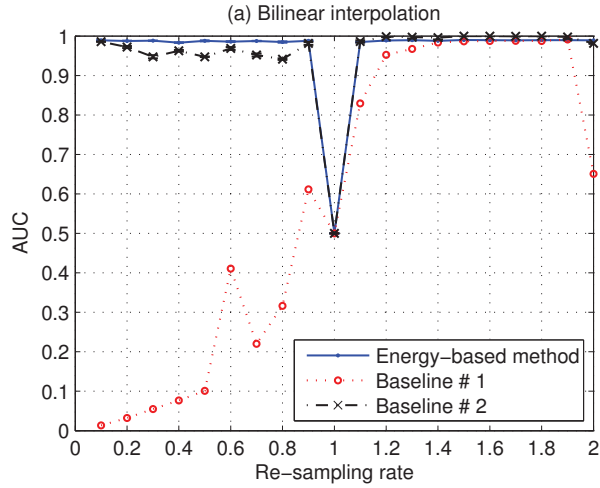


Fig. 4. Comparison of the energy method and the baseline, where one SVM-based classifier is trained for each re-sampling rate. The error bars indicate the standard deviations after 10 trials.

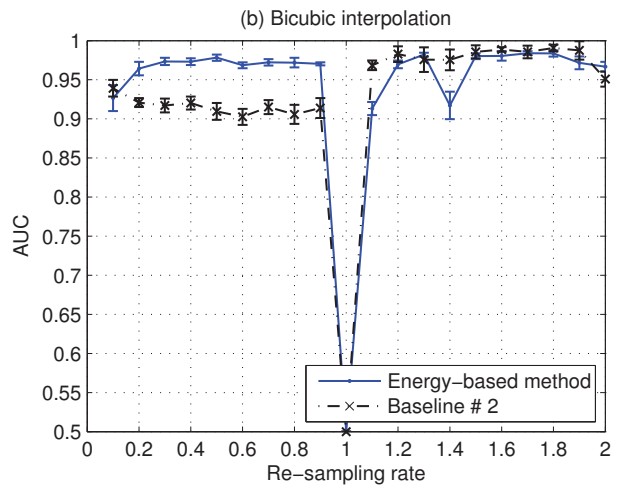
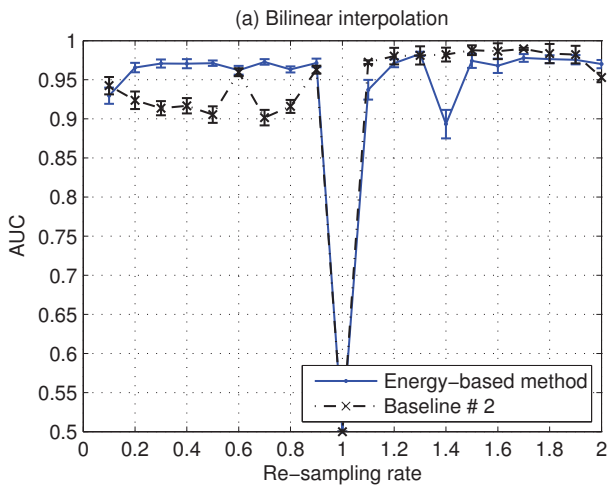


Fig. 5. Comparison of the energy method and the baseline, where only one SVM-based classifier is trained for all re-sampling rates. The error bars indicate the standard deviations after 10 trials.

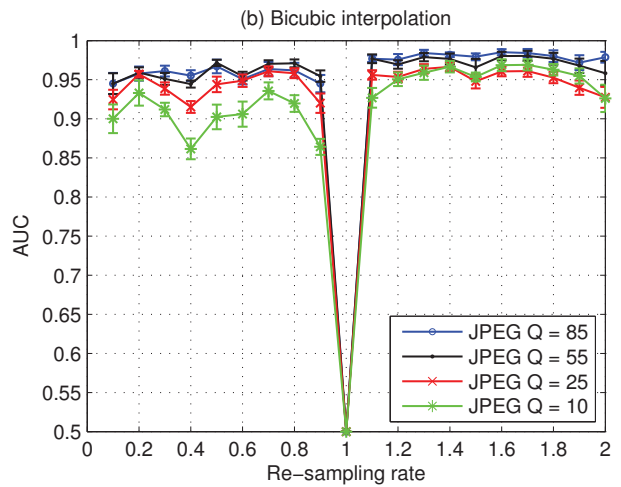
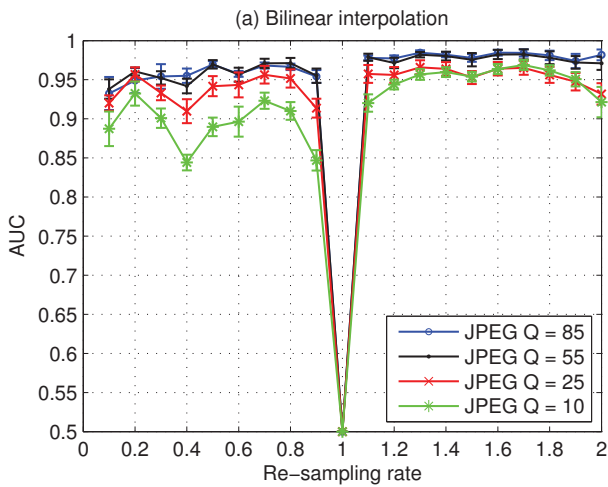


Fig. 6. Detection performance after JPEG comparison for a set of different quality factors based on the energy method, where only one SVM-based classifier is trained for all re-sampling rates. The error bars indicate the standard deviations after 10 trials.

7,500 images from the BOSS database [15] are subsequently compressed with different JPEG quality factors as stated above. Our experiments are then conducted on these JPEG compressed images. We assume that the JPEG quality factor is known. Therefore, for each JPEG quality factor, we train a single classifier for all re-sampling rates based on the normalized energy density approach. Figure 6 shows the experimental results. As expected, the detection performance degrades with decreasing JPEG quality factors.

5. CONCLUSIONS

This paper proposed a new method for detecting re-sampling. It is based on computing the normalized energy density for various window sizes in the second derivative of the frequency domain. The corresponding characteristic curve usually exhibits a peak whose location varies depending on the re-sampling rate. This characteristic curve is exploited to extract a 19-dimensional feature vector that is fed to a SVM-based classifier.

Two forms of SVM classifiers were considered. In the first case, a SVM classifier was trained for each individual re-sampling rate, resulting in 19 classifiers in all. The experimental results showed that the performance of the proposed algorithm is comparable to previous work [9], when the re-sampling rate was greater than 1, and provides superior performance for re-sampling rates less than 1. There was very little difference in performance when either bilinear or bicubic interpolation was used.

In the second case, a single classifier was trained to detect all re-sampling rates. In this case, both our algorithm and the baseline algorithm exhibited a small degradation in performance. However, qualitatively, the relative performance of the two algorithms remained the same, i.e. for re-sampling rates less than 1, our algorithm is almost always superior. For both classifiers, the error bars are small indicating that the algorithm is stable across different re-sampling rates.

Finally, we investigated the detection of re-sampling on JPEG compressed images. Experimental results demonstrated that the detection performance degrades gracefully as the JPEG quality factor is reduced.

In future work, we intend to evaluate the performance of our proposed method with other interpolation techniques such as seam carving and sparse coding. In addition, we will further investigate the impact of various signal post-processing operations on the detection performance.

6. REFERENCES

- [1] H. Farid, "Image forgery detection," *IEEE Signal Processing Magazine*, vol. 26, no. 2, pp. 16–26, 2009.
- [2] I. J. Cox, M. L. Miller, J. A. Bloom, J. Fridrich, and T. Kalker, *Digital Watermarking and Steganography*, 2nd ed. Morgan Kaufmann, 2008.
- [3] A. C. Popescu and H. Farid, "Statistical tools for digital forensics," in *Proceedings of the 6th Information Hiding Workshop, Lecture Notes in Computer Science*, vol. 3200, pp. 128–147, 2004.
- [4] 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, 2006.
- [5] J. Lukáš and J. Fridrich, "Estimation of primary quantization matrix in double compressed JPEG images," in *Proceedings of the Digital Forensic Research Workshop*, 2003.
- [6] T. Pevný and J. Fridrich, "Detection of double-compression for applications in steganography," *IEEE Transactions on Information Forensics and Security*, vol. 3, pp. 247–258, 2008.
- [7] A. C. Popescu and Hany Farid, "Exposing digital forgeries by detecting traces of re-sampling," *IEEE Transactions on Signal Processing*, vol. 53, no. 2, pp. 758–767, 2005.
- [8] B. Mahdian and S. Saic, "On periodic properties of interpolation and their application to image authentication," in *Proceedings of the 3rd International Symposium on Information Assurance and Security*, pp. 439–446, 2007.
- [9] B. Mahdian and S. Saic, "Blind authentication using periodic properties of interpolation," *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 3, pp. 529–538, 2008.
- [10] M. Kirchner, "Fast and reliable resampling detection by spectral analysis of fixed linear predictor residue," in *Proceedings of the 10th ACM Workshop on Multimedia and Security*, pp. 11–20, 2008.
- [11] A. Gallagher, "Detection of linear and cubic interpolation in JPEG compressed images," in *Proceedings of the 2nd Canadian Conference on Computer and Robot Vision*, pp. 65–72, 2005.
- [12] M. Kirchner and T. Gloe, "On resampling detection in recompressed images," in *Proceedings of 1st IEEE International Workshop on Information Forensics and Security*, pp. 21–25, 2009.
- [13] X. Chen, J. Yang, J. Zhang, and A. Waibel, "Automatic detection of signs with affine transformation," in *Proceedings of the Sixth IEEE Workshop on Applications of Computer Vision*, pp. 32–36, 2002.
- [14] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool, "A comparison of affine region detectors," *International Journal of Computer Vision*, vol. 65, 2005.
- [15] "Break our steganographic system," [Online] <http://boss.gipsa-lab.grenoble-inp.fr/>.
- [16] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. Wiley-Interscience, 2000.
- [17] T. Fawcett, "ROC graphs: Notes and practical considerations for researchers," HP Laboratories, 2003.
- [18] F. Uccheddu, A. De Rosa, A. Piva, and M. Barni, "Detection of resampled images: performance analysis and practical challenges," in *Proceedings of the 18th European Signal Processing Conference*, pp. 24–27, 2010.