SPECTRAL METHODS TO DETERMINE THE EXACT SCALING FACTOR OF RESAMPLED DIGITAL IMAGES

Stefan Pfennig
TU Dresden, Faculty of Computer Science
01062 Dresden, Germany
stefan.pfennig@tu-dresden.de

Matthias Kirchner
International Computer Science Institute
Berkeley, CA 94704
kirchner@icsi.berkeley.edu

ABSTRACT
This paper combines analytical models of periodic interpolation artifacts with recent empirical findings on the spectral energy distribution of rescaled images to infer exact transformation parameters in a passive-blind forensic setting. We present a measure to solve a long-known ambiguity between upscaling and downscaling in the forensic analysis of resampled signals and thus substantially limit the range of candidate scaling factors. The effectiveness of our method is backed with empirical evidence on a large set of images and scaling factors.

Index Terms—digital image forensics, resampling detection, transformation parameter, scaling factor

1. INTRODUCTION
The analysis of resampled signals [1–7] has become a standard problem in passive-blind image forensics [8]. Examining traces of geometric transformations (e.g., scaling or rotation) is of particular interest for several reasons. Complex image manipulations often involve such transformations, for instance, when objects have to be aligned in size or perspective. More general, it is beneficial to know as much as possible about an image’s overall processing history, including content-preserving operations such as rescaling of the entire image. Potential applications also include fields like re-synchronization of digital watermarks or camera sensor noise pattern [9].

Most published work concerned with traces of resampling has focused on the mere detection of such post-processing. In particular, periodic interpolation artifacts have attracted wide interest in this context [1–6]. Yet it has also been reported that these periodic artifacts allow inference about the actual geometric transformation the image underwent [1]. Different transformations introduce different periodicities, which are analyzed best in the frequency domain. Here, the position of characteristic resampling peaks hints to parameters of the transformation, for instance the scaling factor [2, 4]. A long-known hurdle in this regard is that there exist classes of scaling factors with exactly the same periodicity, rendering a direct distinction between upscaling and downscaling impossible.

This paper addresses this problem from multiple perspectives. We combine the accuracy of analytical models of periodic interpolation artifacts [4, 5] with recent empirical findings on the spectral energy distribution of rescaled images [7]. The result is a novel approach, capable of solving the ambiguity between upscaling and downscaling and finding exact transformation parameters. Before we present our technique in Sect. 3, Sect. 2 recalls relevant foundations of prior work. Section 4 then reports experimental results. Sect. 5 concludes the paper.

2. FORENSIC ANALYSIS OF RESAMPLED IMAGES

2.1. Linear Predictor Residue
A virtually unavoidable side-effect of resampling is that it introduces periodic linear dependencies between groups of neighboring pixels \(x_{i,j}\). A standard approach to model such interpolation artifacts is in terms of the residual signal \(e_{i,j}\) of a
\( K \times K \) linear predictor \( \alpha [1, 4] \),

\[
e_{i,j} = x_{i,j} - \sum_{(k,l) \in (-K,...,K)^2(0,0)} \alpha_{k,l} x_{i+k,j+l}.
\]

The idea behind this approach is that large absolute prediction errors indicate a minor degree of linear dependence and vice versa. A so-called \( p\)-map as a measure for the strength of linear dependence can be derived from the prediction error, which is modeled as a zero-mean Gaussian random variable.\(^1\)

Figure 1 presents typical detection results for two grayscale images, scaled up to 111 % and 150 % of the original, respectively (left column). The calculated p-maps are displayed in the middle column. Resampling causes periodic artifacts in the p-maps, resulting in distinct peaks in their spectra.

2.2. Energy-Based Analysis

More recently, also spectral signal characteristics have been exploited for resampling detection. Here, the basic assumption is that resampling influences the way the image’s energy is distributed in the frequency domain. While upsampling leads to relatively less energy in high-frequent parts of an image, downsampling has exactly the opposite effect. Denoting \( \tilde{X}_{u,v} \) as the discrete Fourier transformation of a (highpass-filtered) \((2N+1) \times (2N+1)\) image\(^2\) \(\tilde{x}_{i,j}\), this effect can be measured in terms of the normalized energy density \([7]\), \(\tilde{E}_n(z)\),

\[
\tilde{E}_n(z) = \frac{\sum_{(u,v) \in (-zN,...,zN)} |\tilde{X}_{u,v}|^2}{z^2 \cdot \sum_{(u,v) \in (-zN,...,zN)} |\tilde{X}_{u,v}|^2},
\]

for spectral analysis windows of relative size \( z \in (0, 1] \).

Figure 2 illustrates the average shape of the normalized energy density for a number of representative scaling factors and window sizes. The curves were obtained from random \(200 \times 200\) crops of each 1000 rescaled images of the ‘Dresden Image Database’ \([10]\). As expected, the curves indicate a higher concentration of energy in the lower frequencies of upscaled images. Resampled images can be detected, for instance, by measuring \(\tilde{E}_n(z)\) for a number of appropriate window sizes and feeding these values into a classifier \([7]\).

2.3. Inferring Transformation Parameters

While most of the literature has been concerned with the mere detection of resampled signals, it is also possible to infer parameters of the underlying geometric transformation from the image under analysis. Returning to the example in Fig. 1, observe that different scaling factors yield specific periodic artifacts in the p-map. Hence, also the position of distinct resampling peaks in the frequency domain varies.

\(^1\)See \([1, 4]\) for details about computing the p-map.
\(^2\)We assume a square image without loss of generality.
the principle shape of this empirical metric gradually varies as a function of the scaling factor—finding the exact transformation parameter from a single image is infeasible. Moreover, it seems impractical to train classifiers for each single scaling factor. In this sense, energy-based approaches are more suitable to determine a range of potential scaling factors.

3. A COMBINED APPROACH

Both predictor-based and energy-based analyses are in general well able to detect resampled images for a wide range of scaling factors. However, with respect to the distinction of specific scaling factors, both approaches have their own limitations. While there exist analytical yet ambiguous expressions to explain the exact position of characteristic resampling peaks in the p-map’s spectrum, normalized spectral densities can very likely only point to an interval of plausible scaling factors.

Following the above discussion, the strengths and caveats of each individual technique suggest a combined approach. Figure 4 gives an introductory example and illustrates how characteristic resampling peaks and spectral energy distribution may complement each other. More specifically, we adapt Feng et al.’s resampling detection approach [7] and measure approximate scaling factor \( s \approx \arg\min_{z \in S} \left| \frac{1}{n + d}, \frac{1}{n - d} \right| \) of both transformations is the same. However, their spectral densities differ and allow to distinguish between upsampling and downscaling.

Figure 5 puts this combination into a more formal context. For a given image under analysis, the algorithm examines the p-map’s spectrum for distinct peaks. Assuming isotropic scaling, a necessary condition for a spectral component to be considered as potential resampling artifact is that there exist four symmetric peaks on the main axes with the same horizontal/vertical distance \( d \) from the center. The result is a set \( S \) of plausible scaling factors, which can be found from Eq. (3) as

\[
S = \bigcup_{n \in \mathbb{N}} \left\{ \frac{1}{n + d}, \frac{1}{n - d} \right\}. \tag{5}
\]

Note that for the often researched range of scaling factors \( \frac{1}{3} \leq s \leq 2 \), it is sufficient to consider cases with \( n = 1 \). We also note that—in practice—the accuracy of the elements in set \( S \) increases with the available spectral resolution, which itself depends on the size of the image under analysis.

A candidate scaling factor \( \hat{s} \) is chosen from set \( S \) by taking information about the image’s spectral density into account. More specifically, we adapt Feng et al.’s resampling detection approach [7] and measure \( \hat{E}_n(z) \) for a sequence of relative window sizes \( z \). This sequence is interpreted as feature vector that is used to assign the image under analysis to a range of scaling factors with presumably similar energy density characteristics. The result of this (multi-class) classification procedure is an approximate scaling factor \( \hat{s}_E \), representative of the respective class. Typically this scaling factor will be the center of the

![Fig. 3. Characteristic frequency of isotropic rescaling with scaling factor \( s = 1/s \).](image)

![Fig. 4. Resampling artifacts in the spectrum of the p-map (left column) and the energy density distribution (right column) for 150 % (top) and 75 % (bottom) rescaling of the same original image. While characteristic resampling peaks coincide, the images’ spectral densities allow to solve the ambiguity.](image)

![Fig. 5. Combination of periodic resampling artifacts and energy density characteristics to determine the scaling factor.](image)
While we expect to find distinct peaks in the p-map’s spectrum, we found that this has no noteworthy effect, if the number of classes also amplifies the impact of mis-classifications, so that a particular choice of class intervals may affect the accuracy of the combined approach in various ways.

4. EXPERIMENTAL RESULTS

For a quantitative evaluation of the proposed scheme, we use a subset of the ‘Dresden Image Database’ [10]. Our image set comprises overall 600 grayscale images of size 1024 × 1024, which we obtained as random crops from never-compressed, full-resolution RAW images, demosaiced using dcraw with standard settings. If not stated otherwise, each image was rescaled 150 times (using bicubic interpolation) with scaling factors in the range [1/2, 2], sampled in equidistant steps of width Δ = 0.01 and cropped to a size of 512 × 512 after rescaling. We use a random subset of each 100 images per scaling factor to train the classifier. The remaining 150 · 500 images are used for testing. The classifier runs as multi-class SVM with a radial basis function and a one-against-one decision procedure. The p-maps are computed using a preset linear filter kernel of size 3 × 3 [4]. Prior to inspection for distinct peaks, undesired low-frequency components in the p-map’s spectrum are suppressed by normalization with respect to a median-filtered version of the spectrum [5]. Set S is constructed from the strongest peaks on the main axes of the p-map’s spectrum, and estimates ˆs_E are determined from the center of the corresponding class interval.

4.1. Baseline Results

Figure 6 reports the accuracy of the proposed scheme in terms of the average absolute deviation of the estimated scaling factor ˆs from the true scaling factor ˜s. In this experiment, the width of class intervals was set to W = 0.08, resulting in a total of S = 19 classes. For comparison, the figure also depicts the corresponding graph for estimates ˆs_E as obtained from the analysis based on energy densities only.

In general, we expect a smaller number of classes to increase the likelihood of finding the correct interval. Yet relatively few classes also amplify the impact of mis-classifications, so that a particular choice of class intervals may affect the overall accuracy of the combined approach in various ways.
interval borders. The overall average absolute error of the combined approach is 0.058, with an average error considerably below 0.1 for the vast majority of individual scaling factors.

We also conducted experiments with varying numbers of SVM classes to represent energy density characteristics of scaling factor intervals with different granularity. Figure 7 presents results for $S \in \{8, 19, 30\}$ classes, with class intervals of width $W \in \{0.2, 0.08, 0.05\}$. For comparison, the figure also reports results for a setup with three variable-width classes that coincide with the relevant principle intervals $I_l, l \in \{2, 3, 4\}$. The figure confirms that a larger number of classes is beneficial for most of the tested scaling factors. While the overall average absolute error for $S = 8$ is 0.066, accuracy increases with smaller class intervals. Moreover, an average error of 0.1 for the three-class experiment indicates that classification only based on the principal intervals is too coarse for a reliable estimation of rescaling factors. Yet too many classes do not add further accuracy, and the overall average absolute error remains at around 0.058 for all tested settings with $W \leq 0.08$.

Observe that both Fig. 6 and Fig. 7 indicate varying errors across different scaling factors, with estimates from upscaled images being generally more reliable. Figure 8 takes a closer look at the source of errors. More specifically, a first potential source of inaccuracy arises from the detection of incorrect peaks in the p-map’s spectrum. The solid blue curve reports the relative frequency of images, where set $\hat{s}$ indeed contains the true resampling factor. The curve highlights—similar to the literature on resampling detection—that downscaled images exhibit relatively weaker periodic artifacts. Consequently, and in accordance to Fig. 6, estimation errors tend to increase with stronger downscaling. It also worth mentioning that the large error at $s = 1$ (i.e., no resampling) results from the absence of characteristic resampling peaks. Hence, arbitrary peaks may be interpreted as artifact and cause wrong decisions.

Moreover, also the analysis of normalized energy densities may introduce errors, as illustrated in the background of Fig. 8. Here, we depict stacked relative frequencies of energy density mis-classifications, with different colors denoting the distance between class labels of the true and the detected class. The dotted region corresponds to correct classification. This graphical representation indicates that energy-based classification is most reliable at the center of class intervals, but looses accuracy towards class borders. However, most mis-classifications result in relatively low class distances, which are only relevant close to the border between ambiguous rescaling factor intervals. Observe that Fig. 6 supports this hypothesis with relatively larger estimation errors around $s = \frac{2}{3}$ and $s = 1$.

### 4.2. Influence of Image Size and JPEG Post-Compression

Important factors that influence the forensic analysis of resampled images are their size and the strength of JPEG post-compression. Resampling artifacts in small images are generally harder to detect, because fewer samples both increase the impact of noise and reduce the spectral resolution. The former is particularly relevant for the detection of subtle periodic artifacts, whereas a low spectral resolution also affects characteristics of the normalized energy density. Also JPEG post-compression impairs resampling detection. It acts as a low-pass that smooths out subtle artifacts, and it introduces new $8 \times 8$ block periodicities, which have to be ignored by predictor-based methods [2,5].

Figure 9 reports average absolute deviations of the estimated scaling factor for the analysis of each 500 rescaled never-compressed images of sizes $64 \times 64, 128 \times 128, 256 \times 256$, and $512 \times 512$, respectively. As expected, the accuracy generally decreases with smaller image size. Nevertheless, the results also suggest a reasonable accuracy even for (parts of) images as small as $128 \times 128$. Here, the overall average absolute error is 0.074 ($W = 0.08$), i.e., only about two percentage points higher than for the much larger $512 \times 512$ images.

We further repeated the baseline experiment with JPEG post-compression of varying quality factors. Figure 10 depicts the corresponding results for images of size $512 \times 512$ and JPEG qualities between 100 and 90. In general, the graphs of average absolute estimation errors are in line with our expecta-

### Table 1. Overall average absolute deviation of estimates $\hat{s}$ and $\hat{s}_E$ from the true scaling factor for different JPEG qualities. * refers to no compression. Images of size $512 \times 512$, $W = 0.08$.

<table>
<thead>
<tr>
<th>JPEG qual.</th>
<th>estimate $\hat{s}$</th>
<th>JPEG qual.</th>
<th>estimate $\hat{s}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{s}_E$</td>
<td></td>
<td>$\hat{s}_E$</td>
</tr>
<tr>
<td>*</td>
<td>0.054 0.077</td>
<td>98</td>
<td>0.174 0.081</td>
</tr>
<tr>
<td>100</td>
<td>0.061 0.085</td>
<td>95</td>
<td>0.215 0.111</td>
</tr>
<tr>
<td>99</td>
<td>0.129 0.082</td>
<td>90</td>
<td>0.257 0.167</td>
</tr>
</tbody>
</table>
tion. While compression with quality 100 yields only slightly inferior estimates than for uncompressed images, the average accuracy decreases rapidly with lower quality levels. This effect is particularly strong for large scaling factors, where periodic resampling artifacts are of high frequency (cf. Fig. 3).

Table 1 illustrates that the poor performance is mainly due to the well-known vulnerability of predictor-based analyses to JPEG compression. More specifically, the table compares estimation errors of the combined approach and a plain energy-based analysis. It indicates that moving from JPEG quality 100 to 99 has tremendous impact on the former, whereas the latter is more robust and barely affected. The figures even suggest to ignore p-map characteristics for JPEG qualities below 99. Then, however, forensic investigators are again left with the inherent inability of energy density characteristics to determine exact scaling factors (cf. Fig. 6), which emphasizes the general need for further research on resampling artifacts in JPEG post-compressed images.

5. CONCLUDING REMARKS

This paper presented a method to determine the scaling factor from resampled images in a passive-blind forensic setting. Our technique substantially limits the range of candidate scaling factors by combining analytical models of periodic interpolation artifacts with characteristics of the spectral energy distribution of rescaled images. Ambiguities in the position of characteristic resampling peaks are resolved by an energy-based multi-class classification procedure. Experimental results from a large set of images confirm a high accuracy for a wide range of scaling factors. Applications may not only include the analysis and detection of resampled images itself, but also the resynchronization of camera sensor noise pattern [9] or digital watermarks. Future research has to investigate means to extend this work to strong(er) downsampling and JPEG post-compression—processing primitives that are well-known to attenuate periodic resampling artifacts.

6. REFERENCES