BLOCK SIZE FORENSIC ANALYSIS IN DIGITAL IMAGES

Steven Tjoa[*]*, W. Sabrina Lin*[*]*, H. Vicky Zhao*[†]*, and K. J. Ray Liu*[*]

[*]Dept. of ECE, University of Maryland – College Park, MD 20742 USA
[†]Dept. of ECE, University of Alberta – Edmonton, AB TG6 2V4 Canada

ABSTRACT

In non-intrusive forensic analysis, we wish to find information and properties about a piece of data without any reference to the original data prior to processing. An important first step to forensic analysis is the detection and estimation of block processing. Most existing work in block measurement uses strong assumptions on the data related to the block size or the method of compression. In this paper, we propose a new method to estimate the block size in digital images in a blind manner for use in a forensic context. We make no assumptions on the block size or the nature of any previous processing. Our scheme can accurately estimate block sizes in images up to a PSNR of 42 dB where block artifacts are perceptually invisible. We also offer a measure of detection accuracy which correctly classifies an image as block-processed with a probability of 95.0% while keeping the probability of false alarm at 7.4%.

Index Terms— Image and video forensics, image coding, image block size estimation, block artifact analysis

1. INTRODUCTION

Traditional approaches to multimedia security protect content using additive operations. For example, watermarking embeds a signal imperceptibly such that the additive signal is robust and traceable. In order to add the watermark, we require access to the original host signal. However, in many scenarios, we may not even have access to the host signal, and therefore we cannot enforce protection through any extrinsic means. With non-intrusive forensic analysis, the forensic analyst only has access to an output signal in a raw format, without any header information or metadata. Past operations performed upon the signal leave artifacts which become an intrinsic part of the signal, much like a fingerprint. We analyze these artifacts to identify the history of operations.

There are many useful purposes of a non-intrusive forensic system. For example, we often wish to determine the specific encoding mechanism used within a broad category of coders to detect potential *patent infringement*. This service is essential for detecting infringement in software and hardware products that are distributed for profit. By analyzing the artifacts that lossy coders leave behind, we can tell which coder was used along with its parameters. We can also certify the *datapath integrity* of our data. The creation, coding, and delivery of multimedia data constitutes a unique datapath. To ensure that the received data has been processed by the appropriate trusted entities, we must validate the datapath by identifying each of its steps: acquisition, source coding, channel coding, and transmission. We assess the authenticity of the received data by identifying the particular mechanism used in each step of the datapath along with its parameters.

[*]Email: {kiemyang, wylin, kjrliu} @ umd.edu.
[†]Email: vzhao@ece.ualberta.ca.

To even begin forensic analysis for digital images, we must first address the presence of block processing on our image data. For forensic analysis of block-based coding schemes, estimating the *block size* is an obvious and crucial first step, because inaccurate block size estimation can possibly invalidate subsequent forensic tests. For example, suppose we wish to determine the quantization scheme applied upon a single 8-by-8 block along with the quantization parameters without any knowledge of the coding scheme. If we incorrectly estimate that the block size is 16-by-16, then our forensic system will attempt to detect quantization parameters for four 8-by-8 blocks simultaneously, which will result in a faulty analysis.

Block artifact measurement is a well-established research area, with purposes primarily related to image restoration and distortion measurement. However, artifact measurement for the purpose of *forensic analysis* has not been explored. Here, we pose the following question: given a compressed image, can we detect the presence of block processing? If so, can we estimate the block size? Existing work in block artifact measurement is not tailored to answer this question due to strong assumptions placed upon the input data. For example, typical block artifact measurement and reduction methods such as the methods by Minami and Zakhor [1], Liu and Bovik [2], Weerasinghe *et. al.*[3], and Gao *et. al.*[4], all assume a priori that the image data is compressed through an established scheme such as JPEG or MPEG, with a fixed block size of 8-by-8 or 16-by-16. In a forensic scenario, we have no idea of the block size. For example, JPEG2000 has the option of tiling the image using any block size, up to the size of the entire image, before coding. Some fractal and vector quantization coders use block sizes as small as 2-by-2. Rectangular block sizes are also not uncommon. Given this problem, we need a scheme that does not rely on such strong assumptions regarding the block size.

In this paper, we propose a novel scheme to detect the presence of block processing in an image along with the estimation of the block size. This scheme obtains a one-dimensional block artifact signature for both horizontal and vertical dimensions. We estimate the block size by using maximum-likelihood estimation of the period in the block artifact signature. Next, we propose a measure of detection accuracy in a binary hypothesis test where H_0 represents the absence of block processing and H_1 represents the presence of block processing; this measure can acheive a very high probability of detection while keeping the probability of false alarm low.

2. BLOCK SIZE ESTIMATION

Block artifacts appear as artificial discontinuities within an image. By observing the gradient separately along each dimension, and then averaging this data along the orthogonal direction, we find where block differences occur most often.

Let X be the input image of size $M \times N$, and $X(i,j)$ be the luminance value of pixel (i,j), where $i \in \{0,...,M-1\}$ and $j \in$

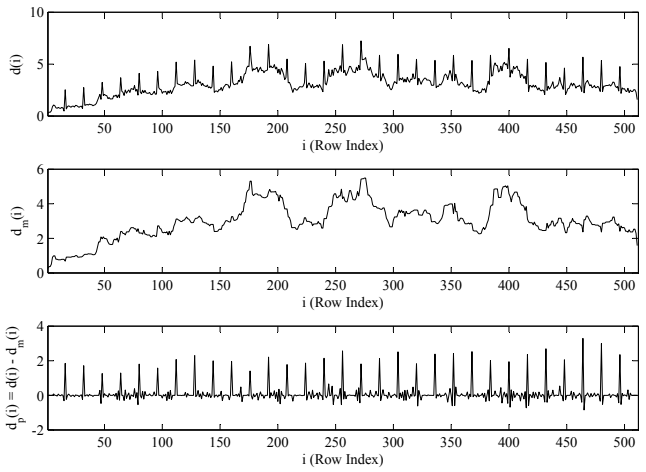

Fig. 1. The signals $\mathbf{d}$, $\mathbf{d}_m$, and $\mathbf{d}_p$ for JPEG-compressed *Lena* with a block size of 16.

$\{0, ..., N-1\}$. The entire following procedure is performed once in the horizontal direction and again in the vertical direction.

First, obtain the gradient image along a desired direction. For example, we operate along the vertical direction shown below.

$$D(i,j) = X(i,j) - X(i-1,j) \qquad i \in \{1, ..., M-1\} \tag{1}$$

Then obtain the average of the gradient magnitude by averaging along the orthogonal direction. For example, we now average along the horizontal direction.

$$d(i) = \frac{1}{N} \sum_{j=0}^{N-1} |D(i,j)| \qquad i \in \{1, ..., M-1\} \tag{2}$$

Let B be the block size, if it exists. If block processing is present, the one-dimensional signal $\mathbf{d}$ will have peaks at multiples of the block size (i.e., at $i = kB$, for $k \in \mathbb{Z}$) superimposed upon a low-frequency signal. To extract these peaks, we use median filtering which eliminates outliers from a signal. In this case, the outliers are the peaks located at $i = kB$ for $k \in \mathbb{Z}$. Let $\mathbf{d}_m$ be the median-filtered version of $\mathbf{d}$. If we subtract $\mathbf{d}_m$ from $\mathbf{d}$ itself, we will obtain the peaks in $\mathbf{d}$ which we call $\mathbf{d}_p$, the one-dimensional block artifact signature.

$$d_m(i) = \text{med}\,\{d(i-1), d(i), d(i+1)\} \tag{3}$$

$$d_p(i) = d(i) - d_m(i) \tag{4}$$

Figure 1 shows the signals $\mathbf{d}$, $\mathbf{d}_m$, and $\mathbf{d}_p$ for the 512-by-512 test image *Lena* which has been JPEG-compressed using a block size of 16. The resulting signal $\mathbf{d}_p$ is approximately periodic. Specifically, we expect $\mathbf{d}_p$ to resemble an impulse train, where the magnitude of the impulses is determined by the strength of the block artifacts, and the period of the impulses is determined by the block size.

One problem we face is the presence of spurious peaks in the signal $\mathbf{d}_p$ as a result of edges from objects in the image. Note that an edge of an object will have the same gradient direction along its entire length. However, the gradient direction of block artifacts will oscillate; in other words, the gradient direction of a block artifact is about as likely to be positive as it is negative. If we were to sum the gradients along a block artifact boundary, the values would cancel out and the sum would be close to zero. However, if we were to sum the gradients along an object edge, the sum would be large in magnitude. Therefore, we perform the following check. Let the signal $c(i)$ be a sum of the gradients, not the gradient magnitudes.

$$c(i) = \frac{1}{N} \sum_{j=0}^{N-1} D(i,j) \qquad i \in \{1, ..., M-1\} \tag{5}$$

A peak in this signal will indicate the presence of an edge. Therefore, we find $\mathbf{c}_p = \mathbf{c} - \mathbf{c}_m$, where $\mathbf{c}_m$ is a median-filtered version of $\mathbf{c}$. Then we set $d_p(i) = 0$ for all i where $c_p(i) > \tau$, for some suitable threshold τ. Our experiments indicate $\tau = 5.0$ to be a threshold which eliminates false peaks while preserving the true peaks.

The periodicity of $\mathbf{d}_p$ allows us to use a maximum-likelihood estimation scheme used in pitch detection [5] to determine the period of the signal $\mathbf{d}_p$. Suppose that $\mathbf{d}_p$ consists of a known periodic signal $\mathbf{s}$ plus zero-mean i.i.d. Gaussian noise.

$$d_p(i) = s(i) + n(i) \qquad i \in \{1, ..., M-1\} \tag{6}$$

Let us express the signal $\mathbf{s}$ as a periodic repetition of a signal $\mathbf{q}$ with period B:

$$s(i) = q(i \bmod B) \tag{7}$$

The conditional probability density function of $\mathbf{d}_p$ is

$$p(\mathbf{d}_p|\mathbf{s}, \sigma^2, B) = \frac{1}{(2\pi\sigma^2)^{\frac{M-1}{2}}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{M-1} (d_p(i) - s(i))^2\right) \tag{8}$$

where σ^2 is the variance of the noise $n(i)$. We should note that the noise signal $\mathbf{n}$ is not exactly Gaussian. Due to the nature of the median filter, we can model $n(i)$ as a mixed random variable where the probability density function has an impulse at $i = 0$ and is Gaussian otherwise. Nevertheless, since $\mathbf{n}$ is approximately Gaussian, we continue to apply maximum-likelihood estimation and ultimately achieve excellent results.

To obtain the maximum-likelihood estimate, we maximize the conditional probability density function $p(\mathbf{d}_p|\mathbf{s}, \sigma^2, B)$ with respect to the signal parameter $\mathbf{s}$, the noise variance σ^2, and the period B. Define the set $\mathcal{I}(i;B) = \{kB + i | k \in \mathbb{Z}\} \cap \{1, ..., M-1\}$. Our ML estimate for the signal $\mathbf{s}$ with respect to B is then

$$\hat{s}(i;B) = \hat{q}(i \bmod B; B) \tag{9}$$

where

$$\hat{q}(i;B) = \frac{1}{|\mathcal{I}(i;B)|} \sum_{l \in \mathcal{I}(i;B)} d_p(l) \tag{10}$$

The variance in the noise $\mathbf{n}$ is estimated as

$$\hat{\sigma}^2(B) = \frac{1}{M-1} \sum_{i=1}^{M-1} (d_p(i) - \hat{s}(i;B))^2 \tag{11}$$

Finally, it can be shown that the estimated period $\hat{B}$ which maximizes $p(\mathbf{d}_p|\mathbf{s}, \sigma^2, B)$ is achieved by minimizing the estimated noise variance $\hat{\sigma}^2(B)$ as a function of B:

$$\hat{B} = \underset{B}{\text{argmin}}\, \hat{\sigma}^2(B) \tag{12}$$

This is our estimate for the block size along one dimension (e.g., the vertical dimension). We repeat the process for the other dimension to obtain the estimate for the block size in both dimensions.

Another problem we face with this scheme is that $\hat{\sigma}^2(B) = \hat{\sigma}^2(kB)$ for all $k \in \mathbb{Z}$. We avoid this problem by using a modified estimate for $\mathbf{q}$:

$$\hat{q}(i;B) = \begin{cases} \frac{1}{|\mathcal{I}(i;B)|} \sum_{l \in \mathcal{I}(i;B)} d_p(l) & i = 0, 1, B-1 \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

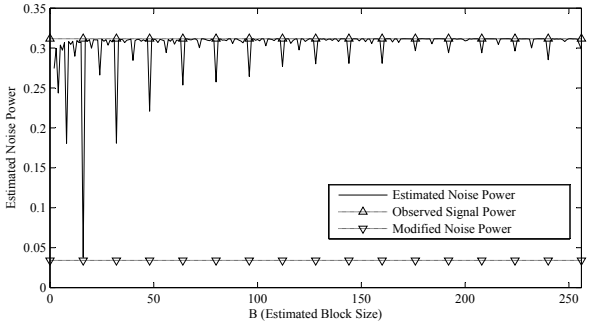

Fig. 2. Noise power $\hat{\sigma}^2(B)$ for JPEG-compressed *Lena* with a block size of 16, the power in the observed signal $\mathbf{d}_p$, and the modified noise power. Each triangle mark corresponds to a multiple of 16.

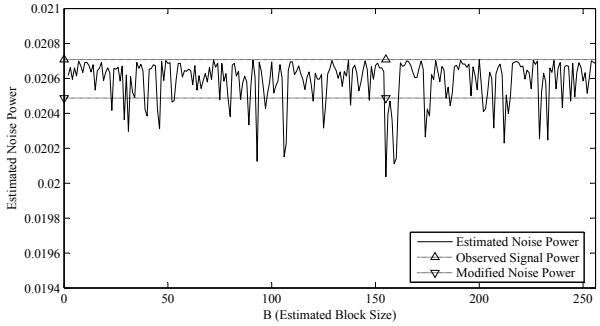

Fig. 3. Noise power $\hat{\sigma}^2(B)$ for uncompressed *Lena*. Note how the observed signal power is lower when block processing is absent.

This choice of $\hat{\mathbf{q}}$ significantly reduces the chance of incorrectly obtaining a multiple of the true block size as our estimate. The plot of $\hat{\sigma}^2(B)$ is shown in Figure 2. Maximum-likelihood estimation easily estimates the correct block size, $\hat{B} = B = 16$.

3. DETECTION ACCURACY

After having found an estimated block size $\hat{B}$, we still ask ourselves how accurate our estimate is, and if block processing is truly present. We can answer this question by considering a simple detection problem with the following two hypotheses:

$$\begin{aligned} H_0: &\quad \mathbf{d}_p = \mathbf{n} \\ H_1: &\quad \mathbf{d}_p = \mathbf{s} + \mathbf{n} \end{aligned} \tag{14}$$

Once again, $\mathbf{d}_p$ is the one-dimensional block artifact signature (i.e., the observed signal). Detection of H_1 implies that our block size estimate is correct, while H_0 corresponds to an incorrect block size estimate or an absence of block processing. Actual execution of a likelihood ratio test requires exact knowledge of $\mathbf{s}$ which we don't have. However, we know that the probability of detection P_D and the probability of false alarm P_F for this test depends on the signal-to-noise ratio in the observed signal $\mathbf{d}_p$. For a high SNR, it is easy to detect the presence of the signal $\mathbf{s}$ (i.e., detect H_1). For a low SNR, $P_D \approx P_F$ for any test, and it is difficult to distinguish between H_0 and H_1. Therefore, we use the signal-to-noise ratio as a measure of our detection accuracy. Figures 2 and 3 illustrate the fact that the observed signal power is much greater when block processing is present versus when it is absent.

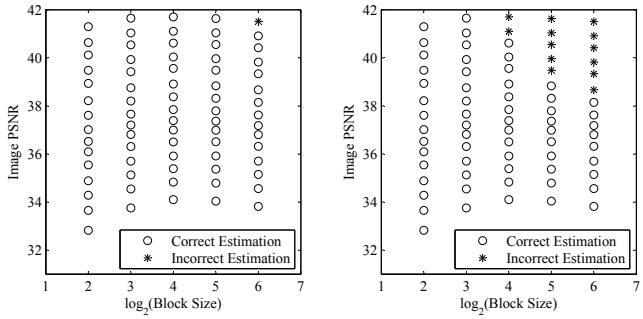

Fig. 4. Block size estimation results for *Lena* in the vertical dimension (left) and the horizontal dimension (right).

Since $\mathbf{s}$ and $\mathbf{n}$ are independent and $\mathbf{n}$ is zero mean, we know that the power in the observed signal $\mathbf{d}_p$ is equal to the power in $\mathbf{s}$, which we call P_s, plus the noise power σ^2. Therefore, the SNR is

$$\text{SNR} = \frac{P_s}{\sigma^2} = \frac{P_{d_p} - \sigma^2}{\sigma^2} \tag{15}$$

Calculating the SNR directly using the signal estimate $\hat{\mathbf{s}}$ could be inaccurate since it relies on the accuracy of the signal estimate itself. In practice, we will use the *observed signal-to-noise ratio* (OSNR) for our measure of detection accuracy

$$\text{OSNR} = \frac{P_{d_p}}{\sigma^2} \tag{16}$$

which we see is clearly related to the SNR.

For our noise variance, we could simply use the value $\hat{\sigma}^2(\hat{B})$ estimated earlier. However, since $\mathbf{s}$ is not exactly periodic, the variance in the peaks of $\mathbf{s}$ will erroneously contribute to the noise variance. As a result, for the purpose of calculating the SNR, we use a modified estimate for our noise variance. Define the set $\tilde{\mathcal{I}}(i;B) = \{1, ..., M-1\} \backslash \{kB + i | k \in \mathbb{Z}\}$. Our modified estimate is

$$\hat{\sigma}^2 = \frac{1}{|\tilde{\mathcal{I}}(0;\hat{B})|} \sum_{i \in \tilde{\mathcal{I}}(0;\hat{B})} (d_p(i) - \hat{s}(i;\hat{B}))^2 \tag{17}$$

In other words, we do not count the variation in the peaks of $\mathbf{s}$ toward the noise variance. The value of the modified noise variance is illustrated in Figure 2 as the bottom line. We observe that the value lies slightly below the minimum variance $\hat{\sigma}^2(\hat{B})$ as expected (where $\hat{B} = B = 16$).

4. RESULTS

Figures 4, 5, and 6 show plots of the block size estimation results for the standard test images *Lena*, *Goldhill*, and *Baboon*. We test for block sizes 4, 8, 16, 32, and 64. To create our block-processed images, we use JPEG compression with quality factors from 20 to 90, though any block processing operation will yield similar results. Each circle represents correct estimation, and each star represents incorrect estimation.

We see that correct estimation varies as a function of both PSNR and block size. Naturally, the strength of block artifacts decreases as image quality increases. As block size increases, the signal $\mathbf{s}$ has fewer periods, and therefore our estimate $\hat{\mathbf{s}}$ is less accurate. For example, as shown in Figure 4, the estimation accuracy for compressed

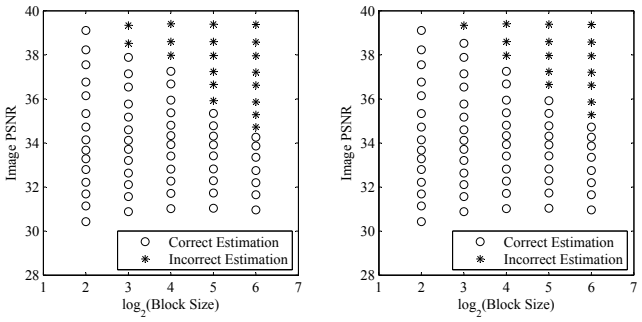

Fig. 5. Block size estimation results for *Goldhill* in the vertical dimension (left) and the horizontal dimension (right).

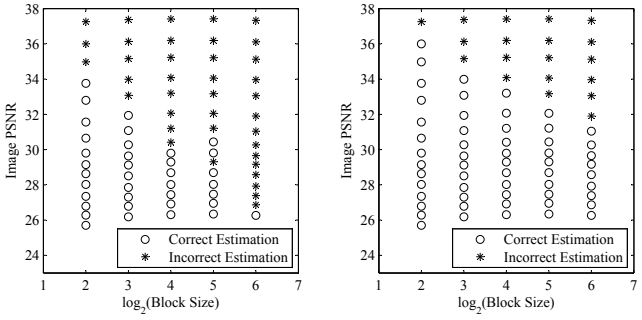

Fig. 6. Block size estimation results for *Baboon* in the vertical dimension (left) and the horizontal dimension (right).

Lena with block sizes of 4 and 8 in the horizontal direction is 100%, while estimation for block sizes of 16, 32, and 64 fails for PSNR above 41.1 dB, 39.5 dB, and 38.6 dB, respectively. Furthermore, estimation accuracy is also data dependent, because high-frequency regions in an image can mask block artifacts. For example, Figure 5 shows slightly worse block size estimation for *Goldhill* than for *Lena*, because *Goldhill* has stronger high-frequency components. Nevertheless, estimation is still accurate at high PSNRs where the artifacts are not perceptually visible.

Furthermore, this scheme is robust under the presence of edges in the image, due to the examination of the signal $\mathbf{c}$, as shown in Eqn. 5. In *Baboon* for the vertical direction, without examining $\mathbf{c}$ to find the edges, the scheme fails due to a spurious edge in the last row of the image, which is most likely an artifact of image acquisition. When we do take into account $\mathbf{c}$, our scheme will ignore this spurious edge, and as a result we obtain a decent estimation accuracy as shown in Figure 6.

We also plot the receiver operating characteristic (ROC) curve to illustrate our detection accuracy as a function of the OSNR threshold in Figure 7 which shows P_D versus P_F for the test in Eq. 14. This plot uses detection results from 24 digital images of natural photographs with varying frequency characteristics, all with size 512-by-512. We test for the same block sizes and quality factors mentioned previously. We see that our scheme can obtain a P_D of 95.0% for a P_F of 7.4%, and a P_D of 98.0% for a P_F of 16.5%. In practice, we can decrease our OSNR threshold to accommodate a higher P_D. The cost of a miss (i.e., detecting no block artifacts when in fact block processing is present) can be significant in a forensic setting where subsequent forensic tests depend on some block size estimate.

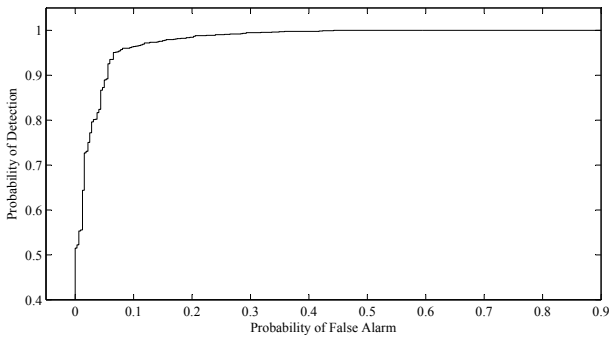

Fig. 7. ROC curve as a function of the OSNR threshold computed from 24 natural photographs.

5. CONCLUSION

We have proposed a block size estimation scheme for use in non-intrusive forensic analysis. Unlike existing block measurement methods, this scheme makes no assumption on the block size or the nature of prior image compression or processing. We reduce our gradient magnitude image to a one-dimensional signal $\mathbf{d}$. Using the fact that gradient magnitudes are greater along block boundaries, we employ median filtering to extract the peaks from $\mathbf{d}$. The block size is then estimated from the extracted signal $\mathbf{d}_p$ using a maximum-likelihood approach. We also propose the use of the observed signal-to-noise ratio (OSNR) as a measure of our detection accuracy. Our scheme works well in high PSNR and for various block sizes (possibly rectangular), and applying a threshold upon the OSNR can accurately detect the presence of block artifacts.

Ultimately, this scheme fits into a broader system for image forensics, particularly one which can identify source coding schemes and parameters. Nevertheless, block size estimation remains a crucial first step in non-intrusive forensic analysis for digital images and video. The necessity of our scheme arises from the fact that, without proper estimation of block processing parameters, subsequent tests on block-processed data would be rendered meaningless.

6. REFERENCES

[1] Shigenobu Minami and Avideh Zakhor, "An optimization approach for removing blocking effects in transform coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 5, no. 2, pp. 74–82, Apr. 1995.

[2] Shizhong Liu and Alan C. Bovik, "Efficient DCT-domain blind measurement and reduction of blocking artifacts," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 12, pp. 1139–1149, Dec. 2002.

[3] Chaminda Weerasinghe, Alan Wee-Chung Liew, and Hong Yan, "Artifact reduction in compressed images based on region homogeneity constraints using the projection onto convex sets algorithm," *IEEE Trans. on Circ. and Sys. for Video Technology*, vol. 12, no. 10, pp. 891–897, Oct. 2002.

[4] Wenfeng Gao, Coskun Mermer, and Yongmin Kim, "A de-blocking algorithm and a blockiness metric for highly compressed images," *IEEE Trans. on Circ. and Sys. for Video Technology*, vol. 12, no. 12, pp. 1150–1159, Dec. 2002.

[5] J. D. Wise, J. R. Caprio, and T. W Parks, "Maximum likelihood pitch estimation," *IEEE Trans. on ASSP*, vol. ASSP-24, no. 5, pp. 418–423, Oct. 1976.