A ROBUST TYPE-III DATA HIDING TECHNIQUE AGAINST CROPPING & RESIZING ATTACKS

Husrev T. Sencar[1], *Mahalingam Ramkumar*[2], *Ali N. Akansu*[1]

[1] New Jersey Institute of Technology
Electrical and Computer Engineering Department
07102 Newark, New Jersey
[2] IDT Corp. / AVWAY.COM Inc.
07102 Newark, New Jersey

ABSTRACT

We propose an oblivious information hiding method which allows watermark recovery for signals subjected to cropping and resampling consecutively. We employ multiple embedding of a watermark signal using a Type - III data hiding method which has a better robustness vs. rate trade off than the conventional methods under mean square error distortion measure. Cyclic autocorrelation features of the cropped-resampled signal are used to estimate the amount of cropping. We analytically showed that within a small error range, it is possible to restore the cropped-resampled signal to the cropped signal. Synchronization of watermark detection in the cropped stego-signal is achieved by designing white noise like watermark signals that are uncorrelated with their shifted replicas. For this purpose we use all-pass filters which are orthogonal to all their cyclic shifts. Freedom to hide data is obtained by modulating the phase of the cyclic all-pass filters. Further, we use Reed-Solomon error correcting codes both for introducing redundancy and achieving synchronization. We also address the issue of data hiding under multiple cropping attacks.

1. INTRODUCTION

Among the three conflicting goals of data hiding, robustness and imperceptibility are more important in watermarking applications rather than hiding rate. Watermarking methods will achieve one degree of robustness as their capability of hiding at relatively higher rates is improved against a variety of well known spatial domain attacks (i.e. cropping, resizing, rotation, uniform/non-uniform scaling, DA-AD conversion, quantization.) However, existence of vast variety of attack scenarios make it a challenge to devise widely accepted or “universal” watermarking schemes.

Spatial domain attacks can be classified into two main groups namely invertible and non-invertible attacks. Invertible attacks can be reversed by some intelligent and usually computationally intense manipulation. Therefore, hiding rate is not decreased. On the other hand, non-invertible attacks like cropping, AD-DA conversion and compression may lead to insignificant hiding rates if they are not taken into account by the designer.

A true watermark embedding methodology should either be invariant to these attacks or include practical means of undoing and reducing the disturbing effects of them. Most of the proposed methods in the literature depend on a particular transform domain for embedding which is immune to some of these attacks. Yet, drastically low hiding rates may still be unavoidable for some others.

In this paper we present an embedding method which allows watermark recovery for signals subjected to cropping and resizing consecutively. These attacks pose a threat of poor watermark detection due to signal transformation and signal loss. We propose practical solutions to aforementioned problems by restoring the resampled signal to its original size and devising ways to synchronize the watermark detection with the embedding. Watermark signals are embedded using a type-III data hiding method, [1, 2], which has a better rate vs. robustness trade off than conventional methods. It is assumed that watermark detector has no access to cover signal (oblivious data hiding.)

Resizing is a transformation onto signal and watermark detection performance relies on either a transform invariant embedding or to invert the transformation before detection. Kutter, [3], proposes a scheme which makes use of autocorrelation to determine affine distortions by comparing the extracted and original watermark locations in an image. We use cyclic autocorrelation peak pattern for computing the amount of cropped data (i.e. number of deleted coefficients in a vector and number of pixels of line in an image.) We analytically show that cropping done up to two different signal locations can be calculated within an ignorable error.

The information loss due to cropping is countervailed by multiple embedding of the watermark signal. Although, multiple embedding is not an ultimate remedy to cropping, the assumption is that at least some replicas of the watermark signal will be undistorted. Erasures in the stego-signal require reinstatement of synchronization. We achieve synchronization by designing watermark signals in form of all-pass filters which are orthogonal to all their cyclic shifts. The phase of the all-pass filter is modulated by the message to be conveyed. Reed-Solomon error correcting codes are used for both introducing redundancy and achieve synchronization as will be evident in section 4.

In the next section, we present a description of the Type-III data hiding method. In section 3, we provide analysis and results for determining the resampling factor (change in aspect ratio for images) in order to restore the stego-signal. Watermark signal design and synchronization issues due to cropping attack are also addressed in the same section. In section 4, we present the results and implementation details for the overall data hiding system.

2. DATA HIDING METHOD

A Typical data hiding system may be represented in an additive channel model as

$$\begin{array}{rcl} \mathcal{W} & : & m \longrightarrow W, \\ \hat{S} & = & \mathcal{E}(S,W) = S + X, \\ Y & = & \mathcal{E}(S,W) + Z = S + X + Z, \\ \hat{W} & = & \mathcal{D}(Y), \\ \mathcal{W}^{-1} & : & \hat{W} \longrightarrow \hat{m}. \end{array} \qquad (1)$$

In the above model m is the message to be hidden, S is the cover data, W is the watermark signal, X is the distortion introduced by the hider, and Z is the intrusion of the attacker. $\mathcal{W}$ is a one-to-one mapping from m to W which transforms message m into a better representation for embedding. Not evident in the model is the distortion constraints imposed on embedding and attacking for keeping the cover signal intact. Ideally, the measure used for quantifying the hider's and attacker's distortion is expected to be in compliance with the perceptual properties of the cover signal. The embedder, $\mathcal{E}$, and the detector, $\mathcal{D}$, may be linear or nonlinear and not necessarily invertible functions. Among various choices of embedder-detector sets $(\mathcal{E}, \mathcal{D})$ available in the literature, data hiding methods may be categorized into three main types.

Type-I methods are very common and simple to implement. Stego-signal is generated by adding the watermark signal to the cover signal. In the model the distortion X is the watermark signal itself. These methods suffer from dramatically low hiding rates because of the non-optimal design which assumes S as a noise and tries to cancel it. Type-I methods are preferable only when the attack is too severe. Besides, they are ideal only for applications for which the cover data is present at the detector.

Type-II methods are characterized by the use of quantizer structures in the embedding and detection. The distortion, X, is a function of S and W. Also, the embedder, $\mathcal{E}$, and detector, $\mathcal{D}$, are inverses of each other. Disadvantage is that the system performs well only if the attack is low. These can also be employed with oblivious data hiding systems at considerable hiding rates.

Type I and II methods have better performances on the two extremes corresponding to severe attack and no attack cases, respectively. An optimal design will be the one that designer has control over the operating characteristics of the method. In [1] Ramkumar proposed a modification to the Type-II methods by removing the invertibility condition on the set $(\mathcal{E}, \mathcal{D})$. In **Type-III** methods added non-invertibility is designed in a particular way that hiding rate is maximized for a presumed attack level. Type-III data hiding is optimal for oblivious data hiding applications.

The embedder being utilized by the data hiding technique is a quantizer characterized by a pair of parameters, period Δ and threshold β where $0 < \beta \leq \Delta$. The form of quantizer used for implementation is a periodic continuous triangular function. Watermark signal to be embedded is limited by the peak values of the periodic function. Embedding is a translation of the input coefficient values by introducing distortions thresholded to $\pm\frac{\beta}{2}$ such that the mapping of embedded coefficient over the periodic function has a minimum Euclidean distance to the watermark signal. The period Δ and threshold β of the quantizer are dictated by the preassigned embedding distortion, above which perceptual features of the cover signal will be considered changed. Among the Δ, β pairs that meet the distortion constraint the one that maximizes hiding rate for a presumed distortion level is picked. Detection of the watermark signal is similar to embedding. The stego-signal is mapped over the periodic function with the same Δ as used for embedding and the fixed threshold $\beta = \Delta$, which adds non-invertibility to the $\mathcal{E}, \mathcal{D}$ set. Mean squared error distance is used as the measure of distortions introduced by information hider and attacker.

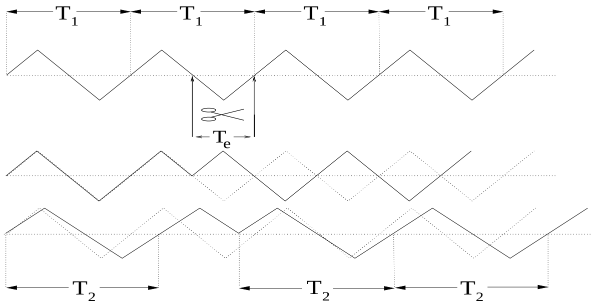


Fig. 1. Representation of cropping and resampling consecutively.

3. CROPPING AND RESAMPLING

Benefits of multiple embedding of the watermark signal is twofold. First, multiple copies of the watermark signal is available for message extraction. Second, autocorrelation techniques can be used to detect the periodicity features of the signal. Figure 1 is a representation of signal cropping and resampling. A signal of length T_1 is repeatedly concatenated to generate a periodic signal (first row.) Then, an amount T_e starting from an arbitrary point of this signal is cropped out (second row) and remaining signal is resampled by some factor $\frac{T_2}{T_1}$ (third row.) Autocorrelation of the resultant signal can be used to estimate T_e with a small error assuming there are un-spoiled periods of signals.

3.1. Autocorrelation for restoring the cropped signal

Consider a signal V, obtained by combining n replicas of the signal W, its cropped version V_C and its cropped-resampled version V_{CR} are generated as displayed in Figure 1. Let signal W be of length T_1, n be a large integer number, T_e be the amount of signal cropped from V, and L be the length of the resultant signal V_{CR}. Then the resampling factor is $\frac{1}{\tau} = \frac{L}{nT_1 - T_e}$. We may also introduce the factor $\epsilon = \frac{L}{nT_1}$ as a measure of deviation from the signal V in terms of size due to resampling.

Autocorrelation function of the signal V_{CR} will point out to existence of two periodic components with the same period, $T_2 = T_1\frac{1}{\tau}$. First component is identified by peaks at every T_2 shift of the origin in the autocorrelation function. On the other hand, second one generates peaks at a shift of $T_2 - T_e\frac{1}{\tau}$ from the zero shift and at every T_2 shift thereafter. The first component is due to existence of the multiple resampled copies of signal W in V_{CR} and second one appears due to erasures. After every shift of $T_2 - T_e\frac{1}{\tau}$ following a T_2 shift the incomplete signal period coincides with a copy of itself and generates a peak in the autocorrelation function. Also, the latter is much more weaker in signal strength compared to the former.

Other than the peak at the zero shift every peak at T_2 shifts will be accompanied by another one assuming n is large enough. Distance d between the peak at kT_2, $k \leq n$, and $(k-1)T_2 + T_2 - T_e\frac{1}{\tau}$ is

$$
\begin{aligned}
d &= kT_2 - \left((k-1)T_2 + T_2 - T_e\frac{1}{\tau}\right), \\
&= T_e\frac{1}{\tau}. \qquad (2)
\end{aligned}
$$

The error in estimating T_e because of scaling by $\frac{1}{\tau}$may also be represented in more explicit terms as

$$
\begin{aligned}
d &= T_e\frac{1}{\tau}, \\
&= T_e\frac{L}{nT_1 - T_e}, \\
&= T_e\frac{\epsilon nT_1}{nT_1 - T_e}, \\
&= T_e\epsilon(1 + \frac{T_e}{nT_1 - T_e}). \qquad (3)
\end{aligned}
$$

In a typical watermark attack scenario the attacker is not willing to make radical changes in the signal size of V. Reevaluating the Eq. 3 for $\epsilon \approx 1$ d can be approximated as

$$d \approx T_e + T_e(\frac{T_e}{nT_1 - T_e}) \qquad (4)$$

The percentage error for using d instead of the actual value of T_e is $\frac{100Te}{nT_1 - Te}$ %. (e.g. For a given vector of length 512 if 25 coefficients are removed and than the cropped vector is resampled back to 512 samples, d is computed as 26.3.) The difference between d and T_e can be minimized by increasing nT_1. When T_e is greater than T_1 an alternative expression of T_e for some integer s and an amount T_{eff} that satisfies $0 < T_{eff} < T_1$ is $T_e = sT_1 + T_{eff}$. The consequence of $T_e \geq T_1$ on autocorrelation function is that s number of peak pairs expected to be observed in autocorrelation function will disappear and remaining peaks will indicate an erasure of T_{eff}. So, effective erasure amount, T_{eff}, is the *modulus* of T_e with respect to T_1. Eq. 3 can be modified to

$$d = T_{eff} + T_{eff}(\frac{T_e}{nT_1 - T_e}) + \frac{sT_1}{\tau}, \qquad (5)$$

where $T_e = sT_1 + T_{eff}$. The first term in Eq. 5 is the distance between the peak pairs and second one is the number of periods of W signals missing in the autocorrelation function of V_{CR}.

3.2. Multiple cropping

Let T_{e1} and T_{e2} be the amounts of the non-overlapping cropped signals from V and $T_{e1} + T_{e2} < T_2$. Autocorrelation function of V_{CR} will have four peaks in every T_2 interval that is $(k-1)T_2$, $k \leq n$, away from zero shift. These peaks will appear at $kT_2 - \frac{T_{e1}+T_{e2}}{\tau}$, $kT_2 - \frac{T_{e1}}{\tau}$, $kT_2 - \frac{T_{e2}}{\tau}$ and kT_2. The last one is due to resampled copies of W and has highest correlation value. Others are due to cropped-resampled copies of W and have smaller values. The distance, d, between the first and last peak in any T_2 interval will be $\frac{T_{e2}+T_{e1}}{\tau}$ where the error in assuming d instead of T_e is as in Eq. 5.

For more numbers of cropping followed by resampling similar analogy is applicable. If $T_{e1}, \ldots, T_{em}$ are the amounts of the non-overlapping cropped signals there may be upto 2^m (2^m peaks only if each cropping is non-overlapping with the others) number of peaks at every T_2 shift of the autocorrelation function. Corresponding peak locations in the autocorrelation function will be at $kT_2 - \sum_{j=1}^{j=m}\frac{T_{ej}}{\tau}$, $kT_2 - \sum_{j=1,j\neq i}^{j=m}\frac{T_{ej}}{\tau}$ for $\forall i$ and $i \leq m$, $kT_2 - \sum_{j=1,j\neq i,l}^{j=m}\frac{T_{ej}}{\tau}$ for $\forall i, l$ such that $i \neq l$ and $i, l \leq m, \ldots,$ $kT_2 - \frac{T_{ej}}{\tau}$ for $\forall j$ and $j \leq m$, and at kT_2 if $\sum_{i=1}^{m} T_{ei} < T_2$. Then, the distance d between the first and last peaks in a T_2 shift can be used as an approximation to the total erasure amount, T_e.

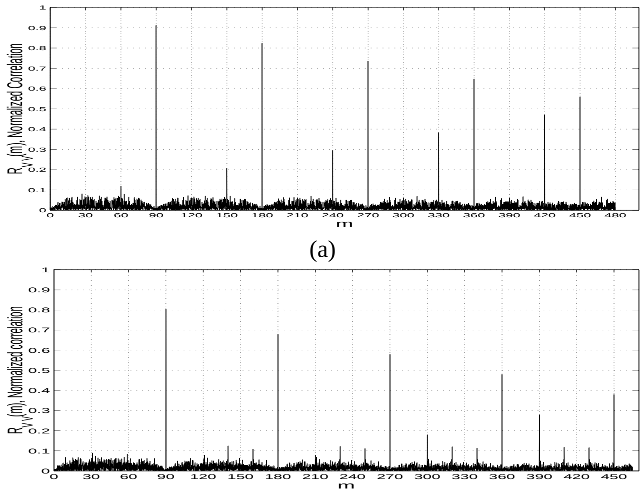
3.3. Practical concerns

The problem is that some correlation peaks may be buried in noise making peak detection unreliable. Using cyclic autocorrelation and designing white noise like W signals are two remedies against it.

3.3.1. Cyclic autocorrelation

Cyclic autocorrelation enhances the correlation peaks while suppressing the ground noise due to signal wrapping in autocorrelation function. Assuming multiple croppings of $T_{e1}, \ldots, T_{em}$ signals in the range $(\frac{nT_2}{2} - \sum_{j=1}^{j=m}\frac{T_{ej}}{2\tau}, nT_2 - \sum_{j=1}^{j=m}\frac{T_{ej}}{\tau}]$ will be flipped and added onto range $(0, \frac{nT_2}{2} - \sum_{j=1}^{j=m}\frac{T_{ej}}{2\tau}]$. The new location of a peak after signal wrapping always coincides with another peak whose location can be found by subtracting from $nT_2 - \sum_{j=1}^{j=m}\frac{T_{ej}}{\tau}$. For instance peaks at $kT_2 - \sum_{j=1}^{j=m}\frac{T_{ej}}{\tau}$, $kT_2 - \frac{T_{ei}}{\tau}$ for $i \leq m$ and kT_2 will be translated to $(n-k)T_2$, $(n-k)T_2 - \sum_{j=1,j\neq i}^{j=m}\frac{T_{ej}}{\tau}$ and $(n-k)T_2 - \sum_{j=1}^{j=m}\frac{T_{ej}}{\tau}$, respectively. Similarly, assuming V_{CR} has been cropped once, by T_e samples, peaks at kT_2 and $kT_2 - \frac{T_e}{\tau}$ will be translated to $(n-k)T_2 - \frac{T_e}{\tau}$ and $(n-k)T_2$ making it easier for peak detection algorithm.

Figure 2 displays the cyclic autocorrelation functions, $R_{V_C V_C}$, of a periodic signal. Signal W is assumed to be of size 90 and V is generated using 11 replicas of it. Figure 2-a is cropped once by removing first 30 samples of sixth period and Figure 2-b is cropped twice by removing middle 40 samples of third period and last 20 samples of fifth period (lower). Every shift of size 90, corresponding to size of W, contains two peaks in 2-a and 4 peaks in 2-b. The distance between the peaks in the former is 30 and the distance between the first and fourth in the latter is 60.



(a)

(b)

Fig. 2. Computing total cropped amounts using cyclic autocorrelation $R_{V_C V_C}$. **(a)** Cropping once, $T_e = 20$. **(b)** Multiple cropping, $T_{e1} = 40$ and $T_{e2} = 20$.

3.3.2. Watermark Signal Design

Watermark signal W in the Eq. 2 is a one-to-one mapping from message m that embedder makes better use with while embedding.

One new constraint is that the signal V should have better autocorrelation features based on the choice of signal W. Designing W as an all-pass filter which is orthogonal to all its cyclic shifts, [4], gives one freedom to hide information by modulating the phase of the W as well as the improved autocorrelation properties. An all-pass filter W of size T_1 has $\frac{T_1-1}{2}$ for T_1 odd ($\frac{T_1-2}{2}$ for T_1 even) degrees of freedom.

3.4. Synchronization

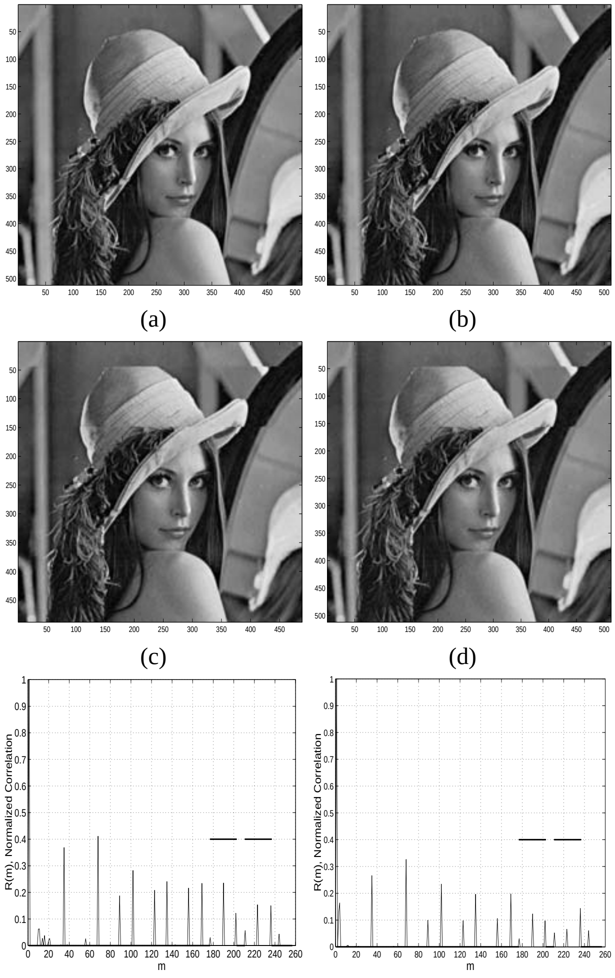
Restored cropped signal must be repartitioned to get signal W back. Since it is not certain which partitions are affected from cropping, extractor needs some markers for re-establishing the synchronization. Most of the partitions will contain signal W or a translated version of it. While some other partitions may have cropped and translated versions of the watermark signal. We use Reed-Solomon error correcting codes for generating watermark signal W and handling synchronization. Given enough redundancy both robustness against signal loss will be achieved and errorless decoding of most of the partitions will be possible at some cyclic shift of the partition.

4. RESULTS

We implemented the methodology on 512×512 graylevel Lena image. Among the various coding strategies we use Hadamard transform matrix and its negated version as the codebook and the orthogonal rows as the codewords that are employed in generating the watermark signal. Message m is assumed to be a sequence of 32 bits that will be conveyed. The message bit sequence is translated into words. Then, the message words are redundancy coded using Reed-Solomon error correcting codes. Encoded message is BPSK modulated and ordered in a way that fulfills the frequency domain symmetry requirements for the phase of the all-pass filter in order to generate the watermark signal W. Watermark signal is chosen to be 32×32 all-pass filter which gives designer $\frac{32\times32-4}{2} = 510$ phase samples to modulate by the coded message m. Then, 16 copies of the watermark signal is embedded throughout the whole image. Distortion introduced by the embedder is approximately 40 dB in PSNR.

Watermarked image is cropped and in order to compensate the reduction in size it is resampled back to its original size. At the extractor a copy of the watermarked, cropped, and resampled image is divided into partitions of size W. Watermark detection for each partition is followed by the two dimensional cyclic autocorrelation of the detected set of signals. Using correlation peak pattern cropped amount is estimated. Extractor, knowing an estimate of the total cropped amount but not their locations, resamples the image back to its size after cropping. So that, disturbing effect of the resampling can be reversed or at least minimized. This image is then partitioned for watermark extraction. Since extracted watermark signal may have been cropped and translated, an immediate detection of message m is not possible. Reed-Solomon codes are used to detect message m from the extracted watermark signal since they are capable of correcting burst error. Two-dimensional signal is shifted in rows and columns until an errorless decoding is possible. High redundancy coding will help detecting message m even under severe signal loss.

Figure 3 displays the results for the described method applied on Lena image, Figure 3-a. Watermarked Lena image is displayed in Figure 3-b where mean squared error per coefficient due to embedding is 9.6. Figure 3-c is the watermarked image cropped in two different locations. Each cropping is 12 lines of pixels in both horizontal and vertical dimensions. Cropped image is resampled back to its original size in 3-d. Figures 3 e-f are the projections of the cyclic autocorrelation function onto horizontal and vertical dimensions. Distance between the first and last peaks in a shift corresponding to a size of watermark signal is 25 which has an estimation error of 1. Image in Figure 3-d is resampled to a size shorter by 25 lines of pixels in each dimension, partitioned in 32×32 blocks and watermark detected. Decoded signals from each block are averaged and watermark detected. Reed-Solomon codes were successful in detecting the 32 bit message m with no errors.



(a) (b)

(c) (d)

(e) (f)

Fig. 3. **(a)** Lena image. **(b)** Watermarked image. **(c)** Cropped image after watermarking. **(d)** Resampled image after cropping. **(e)** Estimation of cropped amounts from the resampled image **(e)** in horizontal dimension, **(f)** in vertical dimension.

5. REFERENCES

[1] Mahalingam Ramkumar and Ali N. Akansu, “Self-noise suppression schemes for blind image steganography,” in *Proc SPIE Multimedia Systems and Applications*, 1999, vol. 3845.

[2] Husrev T. Sencar, Mahalingam Ramkumar, and Ali N. Akansu, “Efficient codebook structures for practical information hiding systems,” in *Proc CISS*, Mar. 2001.

[3] Martin Kutter, “Watermarking resistent to translation, rotation, and scaling,” in *Proc SPIE Multimedia Systems and Applications*, Nov. 1998, vol. 3528, pp. 423–431.

[4] Mahalingam Ramkumar and G. V. Anandand Ali N. Akansu, “On the implementation of the 2-band cyclic filter banks,” in *Proc IEEE International Symposium Circuits and Systems-ISCAS’99*, 1999, vol. 3, pp. 520–523.