OPTIMAL WATERMARK DETECTION ON INTERPOLATED IMAGES UNDER NOISE

Alexia Giannoula, Nikolaos V. Boulgouris, and Dimitrios Hatzinakos

The Edward S. Rogers Sr. Department of Electrical and Computer Engineering,
University of Toronto, Toronto, Ontario M5S 3G4
e-mail: {alexia, nikos, dimitris}@comm.utoronto.ca

ABSTRACT

In this paper, the case where a low-resolution image is initially watermarked and the watermark is afterwards detected on a noisy interpolated version of the watermarked image is investigated, using a correlation detector. Polyphase decomposition is utilized at the detector side in order to enable the flexible formation of a fused image, which is appropriate for watermark detection. The optimal fused correlator, obtained by combining information from different image components, is derived through a statistical analysis of the correlation detector properties and employment of Lagrange multipliers. It is shown that it is always preferable to perform detection on a fused image rather than the original image. Experimental results establish the efficiency of the proposed scheme.

1. INTRODUCTION

Research conducted so far with respect to the copyright protection of digital data through watermarking, establishes the effect of distortions (compression, noise corruption, lowpass filtering) on the system detection reliability. In fact, watermarks have been shown to be severely affected when the host signal goes through a noisy environment.

In many image processing scenarios, it would be desirable to embed a watermark on an image right after acquisition, in order to ensure that no unwatermarked version of the original image is stored or distributed. In the usual case of image resizing, before distribution, at dimensions larger than the original, the watermark information, which was embedded in the low-resolution version upon acquisition, is inevitably spread on the larger image. The larger image could then be compressed and transmitted to any potential recipients.

In this paper, an efficient technique is introduced for exploiting the additional information that may be present on a watermarked image due to distortion. Specifically, linear FIR filters are applied at the detection side, in order to derive several estimates of the original (low-resolution) watermarked image, based on the noisy interpolation samples. By this way, a fused image is generated by combining different filtered components of the interpolated image. A statistical analysis on the properties of the correlation detector, which is applied on the aforementioned fused image (it will be called hereafter *fused correlation detector*), for the case of additive random watermarks, is undertaken and the optimal fusion is contemplated using the Lagrangian approach.

Although correlation-based watermarking techniques [1] are treated, the proposed approach is applicable to most watermarking algorithms. It should be noted that the proposed scheme is not a full-fledged image watermarking technique, but it can be considered an elegant methodology that can be applied *in conjunction* with robust watermarking techniques [2, 3], in order to improve the accuracy of detection.

This work was partially supported by the Bell University Laboratories.

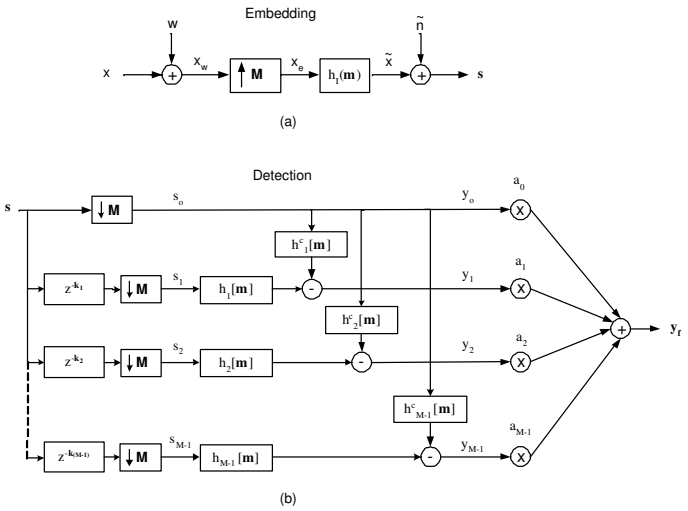


Fig. 1. Block diagram for (a) watermark embedding and (b) fused correlation detection.

2. PROBLEM FORMULATION

Let $x[\boldsymbol{m}]$ be a N_1 x N_2 grayscale image, which will be considered the host signal ($\boldsymbol{m}$ corresponds to the pixels indices and boldface letters represent matrices or vectors). The watermark sequence $w[\boldsymbol{m}]$ of size N_1 x N_2 is an i.i.d. Gaussian-distributed random pattern with zero mean value and variance σ_w^2, generated by a pseudorandom number generator using a suitable key. An additive embedding rule is employed, producing the watermarked image $x_w[\boldsymbol{m}]$:

$$x_w[\boldsymbol{m}] = x[\boldsymbol{m}] + p \cdot w[\boldsymbol{m}] \tag{1}$$

where p is a constant that controls the watermark embedding power.

An expanded image $x_e[\boldsymbol{m}]$ is obtained using an $\boldsymbol{M}$-fold expander:

$$x_e[\boldsymbol{m}] = \begin{cases} x_w[\boldsymbol{M}^{-1}\boldsymbol{m}], & \boldsymbol{m} \in LAT(\boldsymbol{M}) \\ 0, & otherwise \end{cases} \tag{2}$$

where $\boldsymbol{M}$ denotes a 2 x 2 nonsingular integer matrix and $LAT(\boldsymbol{M})$ the lattice defined by $\boldsymbol{M}$ [4]. A linear lowpass filter $h_I(\boldsymbol{m})$ (in-

terpolation filter) is assumed to be applied on $x_e[\boldsymbol{m}]$, generating an interpolated image $\widetilde{x}[\boldsymbol{m}]$. In this manner, the watermark signal, embedded in the low-resolution image, is linearly spread in the new pixel positions of the finer-resolution image $\widetilde{x}[\boldsymbol{m}]$. The eventual high-resolution image $s[\boldsymbol{m}]$ that will be used for detection, is assumed to be a *noisy* version of $\widetilde{x}[\boldsymbol{m}]$. Particularly, it is assumed that the distortion that $\widetilde{x}[\boldsymbol{m}]$ is subjected to, can be modelled as additive noise, i.e:

$$s[\boldsymbol{m}] = \widetilde{x}[\boldsymbol{m}] + \widetilde{n}[\boldsymbol{m}] \tag{3}$$

where $\widetilde{n}[\boldsymbol{m}]$ is the noise signal. The embedding process is schematically described in Figure 1(a).

In the sequel, the detection procedure takes place on the distorted high-resolution image $s[\boldsymbol{m}]$. For this purpose, $s[\boldsymbol{m}]$ is initially decomposed into its constituent polyphase components $s_i(\boldsymbol{m}) = s[\boldsymbol{M}\boldsymbol{m}+\boldsymbol{k}_i],\ \boldsymbol{k}_i \,\epsilon\, \mathcal{L}(\boldsymbol{M}),\ i = 0, \ldots, M-1$, where $\mathcal{L}(\boldsymbol{M})$ denotes the set of all integer vectors of the form $\boldsymbol{M}\boldsymbol{x}$ for $\boldsymbol{x} \,\epsilon\, [0,1)^2$ [4]. The number of vectors $\boldsymbol{k}_i$ is fixed and equal to $M = \mid det\boldsymbol{M} \mid$. The parsing of the samples of $s[\boldsymbol{m}]$ into its M distinct polyphase components $s_i[\boldsymbol{m}]$ is illustrated in Figure 1(b). The polyphase components of the noisy interpolated image $s[\boldsymbol{m}]$ represent different sets of pixels and they cover all the points on the input lattice, where each polyphase component is formed from the points of a coset of that lattice. The zeroth polyphase component $s_0[\boldsymbol{m}]$ of $s[\boldsymbol{m}]$, obtained as shown in Fig. 1(b), essentially represents the original low-resolution watermarked image $x_w[\boldsymbol{m}]$ after noise corruption by a noise signal $n_0[\boldsymbol{m}] = \widetilde{n}_0[\boldsymbol{m}]$ (with variance $\sigma^2_{n_0}$), where $\widetilde{n}_0[\boldsymbol{m}]$ is the zeroth polyphase component of $\widetilde{n}[\boldsymbol{m}]$. The rest of the polyphase components $s_i[\boldsymbol{m}],\ i = 1, \ldots, M-1$, represent interpolated pixels also corrupted by noise.

Our intention is to form M low-resolution images $y_i[\boldsymbol{m}],\ i = 0, \ldots, M-1$, which will be different versions of $x_w[\boldsymbol{m}]$ and whose optimal fusion will provide a new image $y_f[\boldsymbol{m}]$, on which watermark detection will be more reliable than conventional detection on $s_0[\boldsymbol{m}]$. Since $s_0[\boldsymbol{m}]$ by itself is clearly a noisy version of $x_w[\boldsymbol{m}]$, it is used unaltered as $y_0[\boldsymbol{m}]$, i.e:

$$y_0[\boldsymbol{m}] = s_0[\boldsymbol{m}] = x_w[\boldsymbol{m}] + n_0[\boldsymbol{m}] \tag{4}$$

In addition, a sequence of estimates of $x_w[\boldsymbol{m}]$ are derived as follows: the generated polyphase N_1 x N_2 output images $s_i[\boldsymbol{m}],\ i = 1, \ldots, M-1$, are first filtered using appropriate linear filters and then added up to produce

$$y_i[\boldsymbol{m}] = x_w[\boldsymbol{m}] + n_i[\boldsymbol{m}], \qquad i = 1, \ldots, M-1 \tag{5}$$

where $n_i[\boldsymbol{m}]$ is a linear combination of the M noise signals which correspond to the components of $s[\boldsymbol{m}]$. The variance of the noise of each of the above images $y_i[\boldsymbol{m}]$ is $\sigma^2_{n_i}$ and is usually greater than $\sigma^2_{n_0}$ due to the contribution of several noise terms during the derivation of $y_i[\boldsymbol{m}]$.

The proposed technique, subsequently, involves a fusion of the images $y_0[\boldsymbol{m}],\ y_1[\boldsymbol{m}],\ \ldots, y_{M-1}[\boldsymbol{m}]$ by the following rule:

$$y_f[\boldsymbol{m}] = \sum_{i=0}^{M-1} a_i y_i[\boldsymbol{m}] = x_w[\boldsymbol{m}] + \sum_{i=0}^{M-1} a_i n_i[\boldsymbol{m}] \tag{6}$$

where the weight coefficients $a_i,\ i = 0, \ldots, M-1$ are real numbers confined in the $[0, 1]$ interval, which sum up to unity, i.e: $\sum_{i=0}^{M-1} a_i = 1$. The fused image $y_f[\boldsymbol{m}]$ will be finally employed in the correlation detection process described in Section 3. The block diagram of both the embedding and detection procedures can be seen in Fig. 1.

3. STATISTICAL ANALYSIS OF THE CORRELATION DETECTOR

The correlation detector is undertaken in this paper to examine whether a tested image $y_{f,t}[\boldsymbol{m}]$ contains a watermark or not, under a statistical binary hypothesis test, where the following hypotheses are considered:

- H_0: The test image contains the watermark $w[\boldsymbol{m}]$.
- H_1: The test image does not contain the watermark $w[\boldsymbol{m}]$.

Event H_1 occurs either if the test image is not watermarked (event H_{1a}) or if it is watermarked with a different watermark $w_d[\boldsymbol{m}] \neq w[\boldsymbol{m}]$ (event H_{1b}). From equation (6), the three events mentioned above, can be combined in the following expression for the test image:

$$y_{f,t}[\boldsymbol{m}] = x[\boldsymbol{m}] + p \cdot w_e[\boldsymbol{m}] + \sum_{i=0}^{M-1} a_i n_i[\boldsymbol{m}] \tag{7}$$

where the watermark $w[\boldsymbol{m}]$ is indeed embedded in the signal if $p \neq 0$ and $w_e[\boldsymbol{m}] = w[\boldsymbol{m}]$ (event H_0), and it is not embedded in the signal if $p = 0$ (no watermark is present, event H_{1a}) or $p \neq 0$ and $w_e[\boldsymbol{m}] = w_d[\boldsymbol{m}] (\neq w[\boldsymbol{m}])$ (wrong watermark presence, event H_{1b}).

The correlation between the image under investigation $y_{f,t}[\boldsymbol{m}]$ and the watermark sequence $w[\boldsymbol{m}]$ is given by:

$$c_f = \frac{1}{N_1 N_2} \sum_{\mathbf{m}} (x[\boldsymbol{m}]w[\boldsymbol{m}] + pw[\boldsymbol{m}]w_e[\boldsymbol{m}] + w[\boldsymbol{m}] \sum_{i=0}^{M-1} a_i n_i[\boldsymbol{m}]) \tag{8}$$

In order to decide on the valid hypothesis, c_f is compared against a suitably selected threshold T. The performance of such a correlation-based technique can be measured in terms of the probability of false alarm $P_{fa}(T)$ (probability of erroneously detecting the existence of a specific watermark in a signal that is not watermarked or that is watermarked with a different watermark) and the probability of false rejection $P_{fr}(T)$ (probability of erroneously rejecting the existence of a specific watermark in a signal that is indeed watermarked) and can be graphically represented by the receiver operating characteristic (ROC) curve (plot of P_{fa} versus P_{fr}).

For the pseudorandom watermarks employed in this paper, the Central Limit Theorem can be applied, in order to establish that the involved correlator output pdfs under the two hypotheses, $f_{c_f|H_0},\ f_{c_f|H_1}$, attain a Gaussian distribution. Therefore, these pdfs can be described by their mean $\mu_{c_f|H_0},\ \mu_{c_f|H_1}$, and variance values $\sigma^2_{c_f|H_0},\ \sigma^2_{c_f|H_1}$. Using expression (8), the mean $\mu_{c_f} = E\{c_f\}$ and the variance $\sigma^2_{c_f} = E\{c_f^2\} - E^2\{c_f\}$ values of the correlation detector can be evaluated, where $E\{\cdot\}$ denotes statistical expectation. Taking into account the statistical independence between the host signal $x[\boldsymbol{m}]$, the watermarks $w[\boldsymbol{m}],\ w_e[\boldsymbol{m}]$ and the noise signals $n_i[\boldsymbol{m}],\ i = 0, \ldots, M-1$, as well as the fact that the watermark and the noise signals are i.i.d Gaussian-distributed with zero-mean, the corresponding statistics are obtained:

$$\mu_{c_f} \quad = \quad \frac{p}{N_1 N_2} \sum_{\mathbf{m}} E\{w[\boldsymbol{m}]\, w_e[\boldsymbol{m}]\} \tag{9}$$

$$\sigma^2_{c_f} = \frac{1}{N_1^2 N_2^2}\Bigg[\sum_{\mathbf{m}}\Big(E\{x^2[\boldsymbol{m}]\}E\{w^2[\boldsymbol{m}]\} + p^2E\{w^2[\boldsymbol{m}]w_e^2[\boldsymbol{m}]\} + 2p\,E\{x[\boldsymbol{m}]\}E\{w^2[\boldsymbol{m}]w_e[\boldsymbol{m}]\} + E\{w^2[\boldsymbol{m}]\}\sum_i a_i^2 E\{n_i^2[\boldsymbol{m}]\}\Big) + \sum_{\mathbf{m}}\sum_{\mathbf{r},r\neq m} p^2E\{w[\boldsymbol{m}]w[\boldsymbol{r}]w_e[\boldsymbol{m}]w_e[\boldsymbol{r}]\}\Bigg] - \mu^2_{c_f} \quad (10)$$

The above formulas are general and can be applied to all three events, H_0, H_{1a} and H_{1b}. Bearing in mind that $E\{w^2[\boldsymbol{m}]\} = \sigma_w^2$, $E\{w^3[\boldsymbol{m}]\} = 0$ and $E\{w^4[\boldsymbol{m}]\} = 3\sigma_w^4$, for a zero-mean Gaussian watermark $w[\boldsymbol{m}]$ and assuming, also, wide-sense stationarity for the host image $x[\boldsymbol{m}]$, where $\mu_x = E\{x[\boldsymbol{m}]\}$ and $\sigma_x^2 = E\{x^2[\boldsymbol{m}]\} - \mu_x^2$, analytical expressions for μ_{c_f} and $\sigma^2_{c_f}$ can be derived for all three events $H_0,\ H_{1a},\ H_{1b}$:

$$\mu_{c_f} = \begin{cases} p\cdot\sigma_w^2 & , \ if\ w_e = w\ (event\ H_0) \\ 0 & , \ if\ p = 0 \quad (event\ H_{1a}) \\ 0 & , \ if\ w_e = w_d \neq w\ (event\ H_{1b}) \end{cases} \quad (11)$$

$$\sigma^2_{c_f} = \begin{cases} \frac{\sigma_w^2}{N_1N_2}(\mu_x^2 + \sigma_x^2 + 2p^2\sigma_w^2 + \sum_{i=0}^{M-1} a_i^2\sigma^2_{n_i}), & (H_0) \\ \frac{\sigma_w^2}{N_1N_2}(\mu_x^2 + \sigma_x^2 + \sum_{i=0}^{M-1} a_i^2\sigma^2_{n_i}), & (H_{1a}) \\ \frac{\sigma_w^2}{N_1N_2}(\mu_x^2 + \sigma_x^2 + p^2\sigma_w^2 + \sum_{i=0}^{M-1} a_i^2\sigma^2_{n_i}), & (H_{1b}) \end{cases} \quad (12)$$

By observing the above equations, one may easily derive that the mean value of the correlator is the same for both the proposed scenario and the distortion-free scenario. For the special case of $a_0 = 1,\ a_1 = a_2 = \ldots = a_{M-1} = 0$ (which implies watermark detection on the noisy low-resolution image), the correlator mean value for the model under investigation, still, remains the same, while its variance is summarized to: $\sigma^2_{c_f} = \sigma_c^2 + \frac{\sigma_w^2}{N_1N_2}\sigma^2_{n_0}$, where σ_c^2 represents the correlator variance in the noiseless case. In the general case scenario described by equation (12), the fused noise terms are required to be minimized in order to reduce $\sigma^2_{c_f}$ and achieve a gain in the system detection reliability. Therefore, the optimal values $(a_0^*, a_1^*, \ldots, a_{M-1}^*)$, in the sense of minimizing $\sum_{i=0}^{M-1} a_i^2\sigma^2_{n_i}$, need to be determined. *It should be noted that in the noise-free case, the noisy summation terms of equation (12) vanish and therefore, there is no gain, since* $y_0[\boldsymbol{m}], y_1[\boldsymbol{m}], \ldots, y_{M-1}[\boldsymbol{m}]$ *become all identical.*

4. OPTIMAL FUSION USING LAGRANGE MULTIPLIER

By setting $f(a_0, a_1, \ldots, a_{M-1}) = \sum_{i=0}^{M-1} a_i^2\sigma^2_{n_i}$ the optimization problem is equivalent to the minimization of function f, subject to the constraint that

$$g(a_0, a_1, \ldots, a_{M-1}) = \sum_{i=0}^{M-1} a_i - 1 = 0, \ \ i = 0, \ldots, M-1 \quad (13)$$

In order to reach the optimal set of weights $\boldsymbol{a}^* = (a_0^*, a_1^*, \ldots, a_{M-1}^*)$, the *Lagrangian* function for constrained optimization problems [5], is formulated:

$$L = f(a_0, \ldots, a_{M-1}) + \lambda\cdot g(a_0, \ldots, a_{M-1}) \quad (14)$$

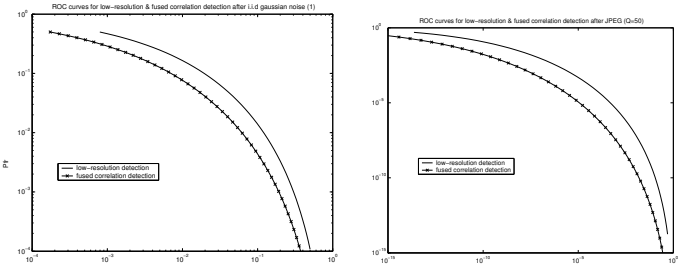
where λ is the Lagrange multiplier. The necessary conditions are:

$$\nabla f(\boldsymbol{a}^*) + \lambda\nabla g(\boldsymbol{a}^*) = \boldsymbol{0}_{Mx1} \quad (15)$$

The above equation yields the solution $a_i = \frac{-\lambda}{2\sigma^2_{n_i}}, \quad i = 0, \ldots, M-1$, which is injected in the constraint condition $g(a_0, \ldots, a_{M-1}) = 0$ to obtain the value of the Lagrange multiplier λ. The solution to the minimization problem is finally derived:

$$\boldsymbol{a}^* = (a_0^*, \ldots, a_{M-1}^*) = \left(\frac{1}{\sigma^2_{n_0}\sum_k \frac{1}{\sigma^2_{n_k}}}, \ldots, \frac{1}{\sigma^2_{n_{M-1}}\sum_k \frac{1}{\sigma^2_{n_k}}}\right) \quad (16)$$

It can be easily verified that $f(\boldsymbol{a}^*)$ corresponds to a global minimum.



(a) (b)

Fig. 2. ROC curves for low-resolution detection and fused-correlation detection after (a) additive i.i.d. gaussian noise with $\sigma_n^2 = 1$ and (b) JPEG compression ($quality = 50$).

5. EXPERIMENTAL RESULTS

A large number of experiments were performed to illustrate the improvement in detection performance of the proposed watermarking scheme. For this purpose, the 256 x 256 grayscale *Lenna* image was watermarked using additive embedding in the pixel domain, producing a watermarked image with Peak-Signal-to-Noise-Ratio (PSNR) approximately equal to 34.2 db. Subsequently, the image was expanded using the matrix: $\boldsymbol{M} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$

In the sequel, the interpolation filter $H_I(z)$[1] was applied in order to generate new image samples. The output was a 512 x 512 image, which was, finally, distorted by additive noise. At the detector side M estimates (where $M = 4$) of the originally available (low-resolution) watermarked image were produced, after the application of the synthesis and cancellation post-filters, $h[\boldsymbol{m}]$ and

[1] with mask: $\{\{1/4, 1/2, 1/4\}, \{1/2, 1, 1/2\}, \{1/4, 1/2, 1/4\}\}$

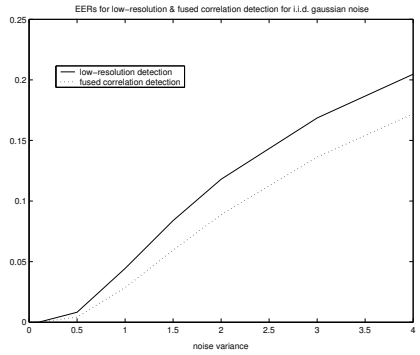


Fig. 3. EER values for low-resolution detection and fused correlation detection after additive i.i.d. gaussian noise corruption, for various noise variances.

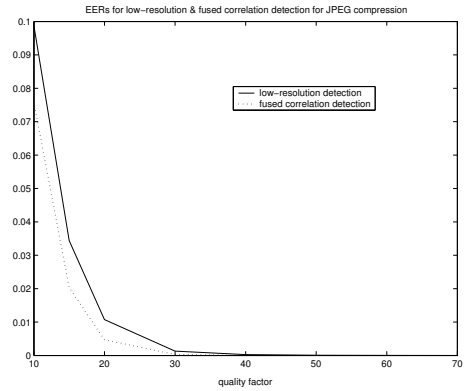


Fig. 4. EER values for low-resolution detection and fused correlation detection after JPEG compression, for various quality factors.

$h^c[\boldsymbol{m}]$. The estimates were fused with the optimal weights and the produced 256 x 256 fused image was finally examined for the presence of a potential watermark using a correlation detector.

All experiments presented in the remainder of this section, were conducted for a total number of 1000 keys and *ROC* curve evaluation was performed under events H_0 and H_{1b}, since the latter represents the worst case assumption (the image being watermarked with a different watermark). In a first set of experiments, additive i.i.d Gaussian noise with zero mean was added to the interpolated image and the watermark detection was performed on the fused image. The EER values (the points on the ROC curve where the probability of false alarm equals the probability of false rejection) for both detection procedures are schematically presented in Fig. 3, where the superiority of the proposed fused correlation-detection scheme is demonstrated for all the values that the noise variance possesses. Calculation of the % EER improvement lead to the conclusion that as larger noise is imposed on the interpolated image, the corresponding % improvement decreases (numerical values are not presented due to lack of space). This implies that the greatest gains are obtained for small noise corruption, which however is usually the case in most watermarking applications, since a potential *attacker* would select a small noise variance to avoid destroying important image information. In all cases, though, the performance is improved by a percentage greater than 15%.

In a typical application scenario, though, the interpolated image will go through one or multiple compression/decompression engines. Therefore, another set of experiments is presented, where the interpolated image was compressed using JPEG of various quality factors (larger quality factors correspond to better image quality). It should be noted that in the case of JPEG compression, the noise signal is not actually uncorrelated, resulting in suboptimal choice of the fusion weights $(a_0, a_1, \ldots, a_{M-1})$. Even in this way, as seen in Fig. 4, there is, still, a considerable improvement in the system detection performance (greater than 22%). Measurement of the % EER improvement in the case of JPEG compression, highlighted the gain in the detection reliability for greater quality factors (weaker compression), which was also seen in the case of i.i.d. gaussian noise corruption. The optimal ROC curve pairs (corresponding to the two alternative detection methods) are shown in Figure 2(a)-(b) for additive gaussian noise and JPEG compression.

6. CONCLUSIONS

An efficient watermark detection technique is introduced in this paper, for improving the correlation-detector reliability on noisy interpolated versions of the originally watermarked image. A fused image was generated by combining different polyphase components of the investigated image. Theoretical analysis was performed for the correlation detector statistics and the optimal fusion was treated using the Lagrangian approach. Significant watermark detection improvements were noted under additive noise corruption and JPEG compression.

7. REFERENCES

[1] J.R. Hernandez and F. Perez-Gonzalez, "Statistical analysis of watermarking schemes for copyright protection of images," *Proceedings of the IEEE*, vol. 87, no. 7, pp. 1142–1166, July 1999.

[2] S. Pereira and T. Pun, "Robust template matching for affine resistant image watermarks," *IEEE Trans. on Image Processing*, vol. 9, no. 6, pp. 1123–1129, June 2000.

[3] S. Stankovic, I. Djurovic, and I. Pitas, "Watermarking in the space/spatial-frequency domain using two-dimensional radon-wigner ditribution," *IEEE Transactions on Image Processing*, vol. 10, pp. 650–658, April 2001.

[4] P. P. Vaidyanathan, "Multirate digital filters, filter banks, polyphase networks, and applications: a tutorial," *Proceeding of the IEEE*, vol. 78, no. 1, pp. 56–93, January 1990.

[5] D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, Inc., New York, 1982.