Chapter 5

Efficient Algorithms in Determining JPEG-Effective Watermark Coefficients

5.1 Introduction

Many digital watermarking schemes have been proposed for copyright protection, data hiding and other purposes. In our previous work as described in Chapter 4, we focus on the tradeoffs between the achievable watermarking data payload, allowable distortion for information hiding, and robustness against attacks. Although many methods have been developed to improve the watermark data payload and robustness while maintaining reliable detection and visual fidelity [80]-[82], few researchers have proposed techniques to identify the exact coefficient locations for watermarking. Thus, we suggested a generic approach for selecting the most effective coefficients for watermark embedding. Using this set of coefficients improves the watermark robustness and reliability while it maintains the watermark visual transparency. To a certain extent, we try to find the performance limit of invisible watermarking for a given natural image under the assumptions of known attack and non-blind detection for DCT-domain watermarking. The non-blind detection can be used in applications such as transaction-tracking. The synchronization attack is not considered as a problem due to non-blind detection. Since digital images are often compressed for efficient storage and transmission, we use JPEG and JPEG2000 as the examples of attacking sources in the design phase.

Although the coefficient selection procedure performs rather well, its computational complexity is very high. Therefore, in this paper, we develop a fast algorithm with nearly no performance loss. In this chapter, the simplified rules for JPEG compression attacking source is presented. Note that the methodology of the coefficient selection procedure in Chapter 4 and the simplified algorithm proposed in this paper both can easily be extended to the other types of attacks. Section 5.2 briefly describes our previous work – theory-based optimal coefficient selection. Section 5.3 de-scribes the newly proposed coefficient selection rules. Simulation results are summarized in Section 5.4 and Section 5.5 concludes this presentation.

5.2 Our Previous Algorithm

Two optimization stages are proposed in Chapter 4 for selecting effective coefficients. One is the robust and imperceptible coefficient selection stage (Stage One), and the other is the detection reliability improvement stage (Stage Two). Stage One conducts a deterministic analysis on the transform coefficients, and then the proper coefficients and the associated watermark strength are determined so that the coefficients after a specified attack can still bear the valid marks. The additive embedding $x'[i] = x[i] + \alpha[i] \cdot w[i]$ is adopted in the DCT domain, where $\alpha[i]$ is the watermark strength of the ith AC coefficient $x[i]$ and $w[i]$ is the watermark bit. All AC coefficients are watermarked. For an attack in either the spatial or other transform domains, the watermarked image is converted back to the spatial domain and the attack is applied. We decode the watermark bits in the DCT-domain. Several different watermark patterns are tested. If all watermark bits associated with a certain DCT coefficient are correctly decoded, this coefficient is retained in the Stage One candidate set. We examine the all-positive and all-negative watermark patterns. When the attack is not applied to individual coefficients in the DCT-domain, we also test the alternate polarity pattern in which the odd-index watermark bits (in zigzag scan order) are +1 and the even-index ones are -1. This is because the attack distortion on a DCT coefficient also depends on its neighboring watermark bits. Our experiments indicated that we can identify robust coefficients with rather high probability by only 4 patterns. The Watson's visual model is adopted for contrast masking threshold computation and the parameter values are taken from the Checkmark package [70].

Some robust coefficients may produce higher detection error probability. Thus, Stage Two calculates the statistical measures on images and attacks, and it discards the weak coefficients. An iterative procedure is proposed and only one coefficient is discarded in each iteration. At the beginning of one iteration, if N coefficients remain, N candidate sets are formed by deleting one coefficient alternatively in this N-coefficient set. That is, there are N-1 coefficients in each candidate set. Then, the watermark detection statistics based on signal dependent channel distortion model [74] and the Bayes' decision rule for each candidate set is calculated for each candidate set. The error detection probability is the average of the false positive probability and false negative probability. Then, the set with the lowest detection error probability is chosen if the average error probability decreases from the previous iteration. The coefficient discarding process is repeated until the overall error probability cannot be further reduced. If there are N selected coefficients at the beginning of Stage Two, and K dropped coefficients in the process, the execution time of Stage Two will be $\mathrm{O}(KN^2)$. Thus, a fast algorithm is very desirable.

5.3 Efficient Robust and Reliable Coefficient Selection Rules

Our goal is finding simplified rules to separate the selected coefficients and dropped coefficients for a given input image based on the theoretically optimized data set derived in

Chapter 4. We adopt a parametric linear classifier for classification [83]. For a parametric classifier approach, the mathematical form of the classifier is specified while a finite set of parameters are left to be determined. These parameters may consist of the expected vectors and covariance matrices. Although linear classifiers are not optimum in some cases, we employ it due to its simplicity.

The general form of a linear classifier (linear discriminant function) regardless of the given input distribution is

$$h(X) = V^T X + v_0 \lessgtr 0 \tag{5.1}$$

where $h(\boldsymbol{X})$ is a linear function of $\boldsymbol{X}$, $\boldsymbol{X}$ is the given input data vector which distributions are not limited, $V = [v_1 v_2 \cdots]^T$ is the coefficient vector, and v_0 is a threshold value. Our goal is to find the optimal V and v_0 for a given distribution. For this, different design criterion $g(\eta_1, \eta_2, \sigma_1^2, \sigma_2^2)$ may be used, and,

$$\eta_i = E\{h(X) \mid \omega_i\} = V^T E\{X \mid \omega_i\} + v_0 = V^T M_i + v_0 , \tag{5.2}$$

$$\sigma_i = Var\{h(X) \mid \omega_i\} = V^T E\{(X - M_i)(X - M_i)^T \mid \omega_i\} V = V^T \Sigma_i V . \tag{5.3}$$

where η_i, and σ_i are the expected value of $h(\boldsymbol{X})$ and variance of $h(\boldsymbol{X})$ for class ω_i, respectively, Σ_i and M_i are the covariance matrix and expected vector M_i for the given input X, respectively.

For any criterion, the optimal V is achieved by maximizing or minimizing the criterion $g(\eta_1, \eta_2, \sigma_1^2, \sigma_2^2)$. And thus,

$$V = [s\Sigma_1 + (1-s)\Sigma_2]^{-1}(M_2 - M_1) , \tag{5.4}$$

where

$$s = \frac{\partial g / \partial \sigma_1^2}{\partial g / \partial \sigma_1^2 + \partial g / \partial \sigma_2^2} . \tag{5.5}$$

The optimal v_0 is accordingly generated by solving

$$\frac{\partial g}{\partial \eta_1} + \frac{\partial g}{\partial \eta_2} = 0 \tag{5.6}$$

Here, we adopt the criterion [83] which measures the between-class scatter normalized by the within-class scatter,

$$g = \frac{P_1\eta_1^2 + P_2\eta_2^2}{P_1\sigma_1^2 + P_2\sigma_2^2}, \tag{5.7}$$

where P_i is the priori probability for class *i*. After generating $\frac{\partial g}{\partial \sigma_i^2}$ and *s* in (5.5), we get optimal V and v_0 by (5.4) and (5.6)

$$V = [P_1\Sigma_1 + P_2\Sigma_2]^{-1}(M_2 - M_1) \tag{5.8}$$

and

$$v_0 = -V^T[P_1M_1 + P_2M_2]. \tag{5.9}$$

The well known Fisher criterion is not adopted since the optimal v_0 cannot be determined through the use of it.

Fig. 5.1. The thirty 256×256 natural images used for rule finding.

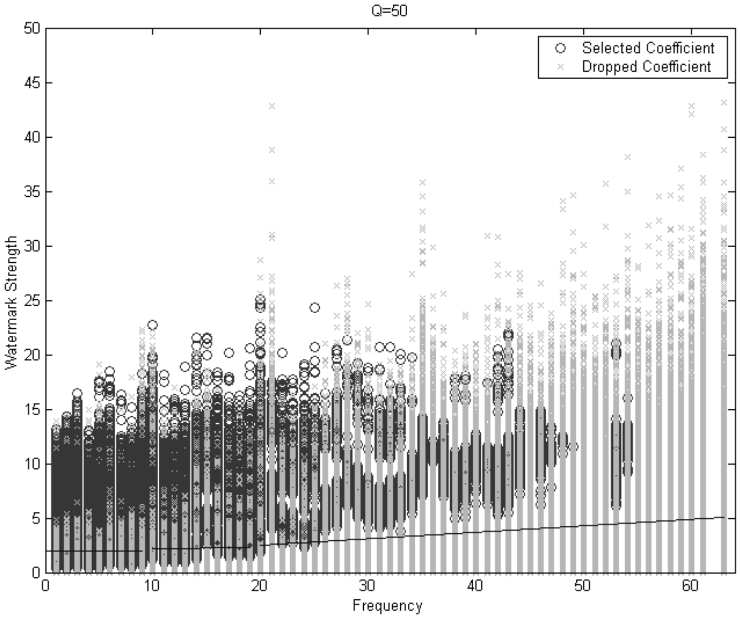
The features in our problem are frequency *f*, amplitude *x* and admissible watermark strength α. Our target is to find a piece-wise linear classifier (discriminator) that separate the selected coefficients from the dropped ones. Thirty natural images are used for rule finding as shown in Fig. 5.1. We have looked at the case that uses all three features (f,x,α) (3-D domain). To simplify calculations, we also search for a 2-D feature space with smallest average misclassification rate. Our experiments show that the "optimal" average misclassification rate in the 2-D space "f vs. α" is only 1% lower than that of the 3-D domain classifier. There are three 2-D domain candidates: (f,x), (f,α), and (x,α). Let S_{fx} , $S_{f\alpha}$ and $S_{x\alpha}$ be the misclassification rate due to the selected coefficients being misclassified as dropped coefficients in the aforementioned three candidate spaces, respectively, D_{fx} and $D_{f\alpha}$, and $D_{x\alpha}$ be the misclassification rate due to the dropped coefficients being misclassified as selected coefficients. According to our experiments, to decrease S_{fx}, $S_{f\alpha}$ and $S_{x\alpha}$, we set $P_1 = 0.4$ and $P_2 = 0.6$. For further improving the classification accuracy, we divide the entire range of a space into three segments, and design one linear classifier for each segment (subspace). The segment partition is done manually based on experience. For (f,x) and (f,α) spaces, the separation is based on f=0~9, f=10~19 and f=20~63. For space (x,α), they are x=0~49, x=50~99 and x=100~∞. Our image data base contains 30 natural images. The training set is generated using the method described in Sect. 2. Four JPEG quality factors ranging from 50 to 80 are used. We adopt the definition of JPEG quantization step size defined in [78]. The misclassification rates in all cases (2D domains) are listed in Table 5.1. Because the best 2-D (f,α) space is 1% worse than the 3-D (f,x,α) classifier, the former is adopted for a much lower computation complexity.

Table 5.1. Misclassification rates in three 2-D feature spaces.

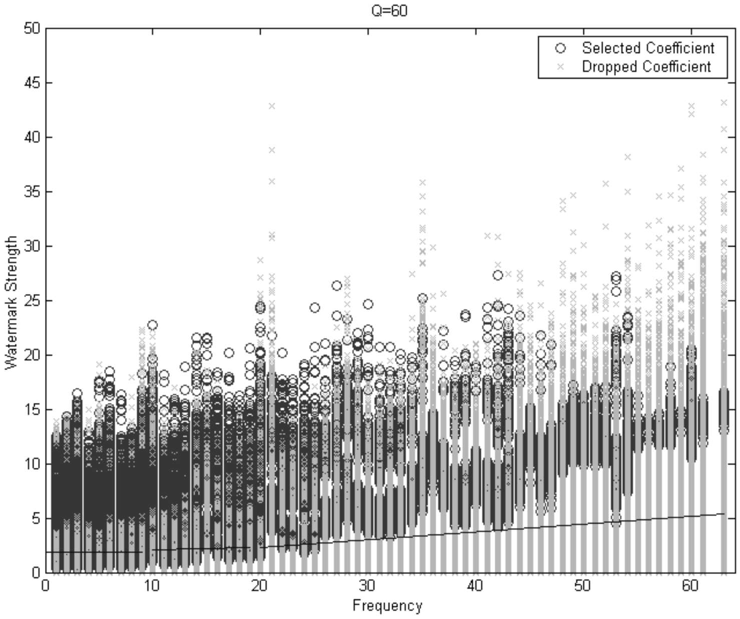
JPEG Quality Factor in Design Phase	S_{fx}	$S_{f\alpha}$	$S_{x\alpha}$	D_{fx}	$D_{f\alpha}$	$D_{x\alpha}$
50	0.31	**0.18**	0.66	0.07	**0.10**	0.40
60	0.29	**0.18**	0.65	0.06	**0.09**	0.38
70	0.27	**0.18**	0.63	0.06	**0.09**	0.35
80	0.27	**0.16**	0.57	0.05	**0.08**	0.30

Table 5.2. The classifiers at different JPEG quality factors.

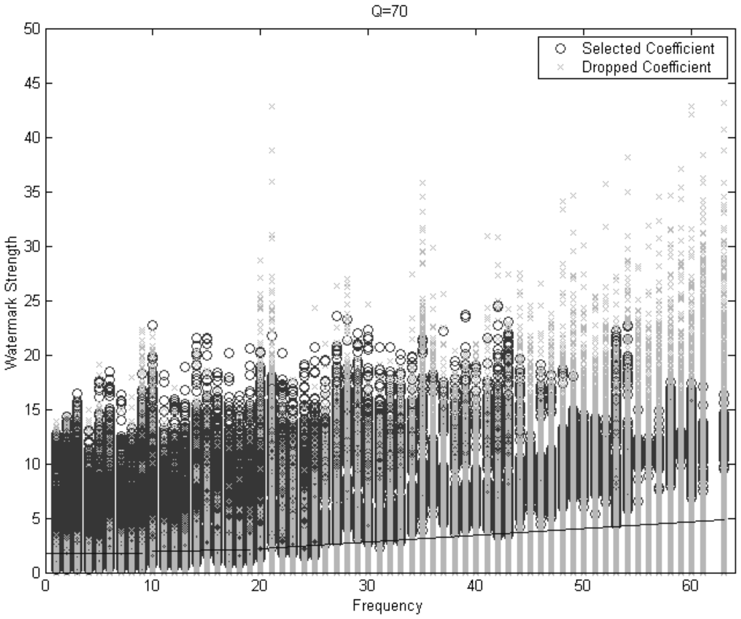
JPEG Quality Factor in Design Phase / Frequency Range	Classifier
Q=50, *f*=0~9	$h(X) = f - 361.56\alpha + 717.86$
Q=50, *f*=10~19	$h(X) = f - 80.42\alpha + 167.63$
Q=50, *f*=20~63	$h(X) = f - 17.18\alpha + 23.86$
Q=60, *f*=0~9	$h(X) = f - 778.65\alpha + 1454.52$
Q=60, *f*=10~19	$h(X) = f - 57.35\alpha + 109.23$
Q=60, *f*=20~63	$h(X) = f - 13.65\alpha - 10.78$
Q=70, *f*=0~9	$h(X) = f - 429.85\alpha + 734.75$
Q=70, *f*=10~19	$h(X) = f - 52.25\alpha + 90.53$
Q=70, *f*=20~63	$h(X) = f - 15.79\alpha + 13.98$
Q=80, *f*=0~9	$h(X) = f + 1885.21\alpha - 3002.29$
Q=80, *f*=10~19	$h(X) = f - 47.28\alpha + 70.24$
Q=80, *f*=20~63	$h(X) = f - 16.65\alpha - 12.02$


Q=50
Selected Coefficient
Dropped Coefficient
Watermark Strength
Frequency
0
5
10
15
20
25
30
35
40
45
50
0
10
20
30
40
50
60

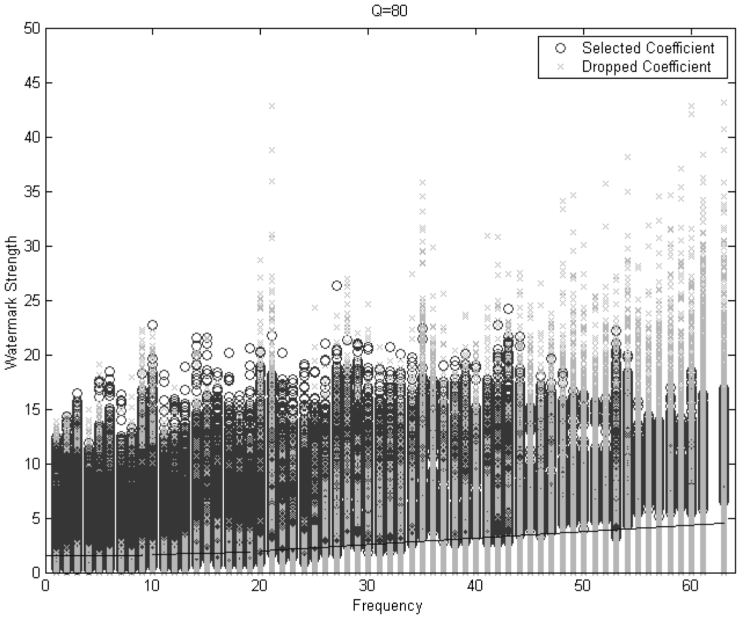
(a)


Q=60
Selected Coefficient
Dropped Coefficient
Watermark Strength
Frequency
0
5
10
15
20
25
30
35
40
45
50
0
10
20
30
40
50
60

(b)



(c)



(d)

Fig. 5.2. The classifiers corresponding to JPEG compression quality factor (a) 50 (b) 60 (c) 70 (d) 80 with coefficients from 30 natural images.

Thus, we can now select effective watermarking coefficients with the simplified rules.

Table 5.2 shows the classifiers (coefficient selecting rules) corresponding to different JPEG quality factor ranging from 50 to 80 in the design phase and Figure 5.2 visualize them. Although these rules eliminate a number of poor candi-date coefficients, the remaining coefficients do not necessarily have the required robustness. Therefore, we apply the original Stage One process to the retained coefficients for further removing weak coefficients.

5.4 Simulation Results

To examine the performance of the proposed rules, we test images which are not used in training. Limited by space, only the results for pictures Lena and Baboon are included. For the JPEG quality factor 50 in the design phase, the PSNR values between the original and the watermarked images are 45.2 dB and 39.98 dB for Lena and Baboon, respectively. And, they are 42.9 dB and 36.82 dB for JPEG quality factor 80 in the design phase. The embedded watermarks are invisible as we inspect them visually.

The comparisons between the original and the simplified schemes are shown in Tables 5.3 and 5.4. Let the overlapped percentage be the number of coefficients selected by both the original Stage One and the simplified scheme divided by the number of selected coefficients by the original Stage One. We find that the overlapped percentage is higher than 70%. The detection error probability using the simplified scheme is still very small (all less than 10^{-135} for Lena). Practically these rules are as good as the original massive iteration scheme. In the case of Baboon image, the overlapped percentage is over 85% and the detection error probability is all less than 10^{-245}.

The data shown in Figs. 5.3, 5.5(a) and 5.6 are each averaged over 5000 watermarked images with different random watermark sequences. Also, the same 5000 watermark sequences are correlated with the unmarked but JPEG compressed image and the results are averaged in Figs. 5.4, 5.5(b) and 5.7. Figure 5.5(a) shows that the selected coefficient survives

JPEG compression at higher quality factors may not survive JPEG compression at lower quality factors. To verify the designed false negative and positive error probabilities, the mean, variance, minimum and maximum values of the normalized correlation sum after the JPEG attacks are computed (The normalization is normalized against the embedded watermark power as discussed in Chapter 4, and thus is not bounded to [-1, 1].) The mean value of the normalized correlation sum C is computed by

$$C = \frac{1}{M}\sum_{i=1}^{M} c[i] = \frac{1}{M}\sum_{i=1}^{M} \frac{y[i]\times(w[i]\times\alpha[i])}{\sigma_d^2} \tag{5}$$

where $y[i]$ is the difference between the DCT coefficients of the received image and the original image, $w[i]$ is the watermark signature and M is the number of selected coefficients. For a watermark sequence, C is compared against the detection threshold which is approximately the average of the mean values of the normalized correlation sum of the watermarked $\mathrm{E}\{c\,|\,\mathrm{H}_1\}$ and unmarked images $\mathrm{E}\{c\,|\,\mathrm{H}_0\}$ as shown in Chapter 4. The presence of the watermark is declared if H1 is favored. In all cases, there is no failure for either watermarked or unmarked 5000 images. Finally, small variance implies lower error detection probability. The variance values $\mathrm{Var}\{c\,|\,\mathrm{H}_1\}$ and $\mathrm{Var}\{c\,|\,\mathrm{H}_0\}$ are all smaller than 0.0018 after JPEG attacks with different quality factors for both watermarked and un-marked cases as shown in Figs. 5.3(b) and 5.4(b). We also test the JPEG-robust watermark against other signal processing attacks as shown in Fig. 5.8 and the data are obtained by averaging over 100 different random watermark sequences. The $\mathrm{E}\{c\,|\,\mathrm{H}_1\}$ is over 0.8 after JPEG2000 attacks at bit rates 0.125 bpp and 0.0625 bpp. We also compare the computational complexity between the original and the simplified stages as shown in Table 5.4. The computational complexity is expressed by the number of processed DCT coefficients. For image Lena at JPEG quality factor 80, the simplified scheme requires roughly 1/266 of the computations of the original scheme (Stage One + Stage Two) for large candidate sets. The simplified scheme does greatly reduce the computational complexity.

5.5 Summary

In this chapter, we propose an efficient algorithm for selecting JPEG-effective watermark coefficients. In most cases, the new scheme uses only 1/100 of the computation needed in the original scheme in Chapter 4. The methodology of both the original coefficient selection procedure in Chapter 4 and the simplified algorithm proposed here can be easily extended to the other types of attacks.

Table 5.3. The comparisons of the selected coefficients for Lena. (The number of selected coefficients by simplified scheme equals to the number of overlapped coefficients by both the original Stage One and simplified scheme.)

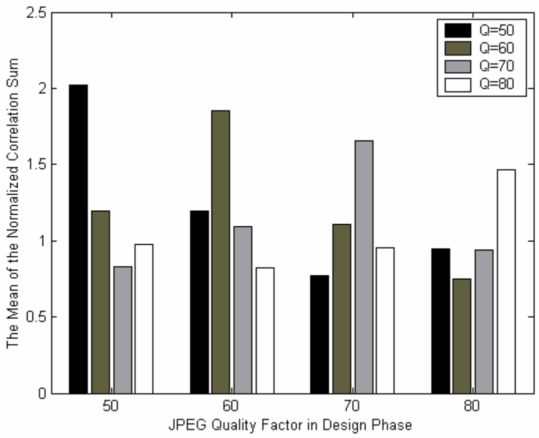
JPEG	Original Scheme			Simplified Scheme	
Quality Factor in Design Phase	No. of Selected Coeff. by Stage 1	No. of Selected Coeff. after Stage 2	Estimated P_{error} after Stage 2	No. of Selected Coeff.	Estimated P_{error}
50	4738	4019	5.5052e-299	3609	2.0978e-136
60	6007	5082	0.0000e+000	4516	6.8039e-181
70	8041	6587	0.0000e+000	5911	2.3203e-253
80	111473	9439	0.0000e+000	8166	0.0000e+000

Table 5.4. The comparisons of the selected coefficients for Baboon. (The number of selected coefficients by simplified scheme equals to the number of overlapped coefficients by both the original Stage One and simplified scheme.)

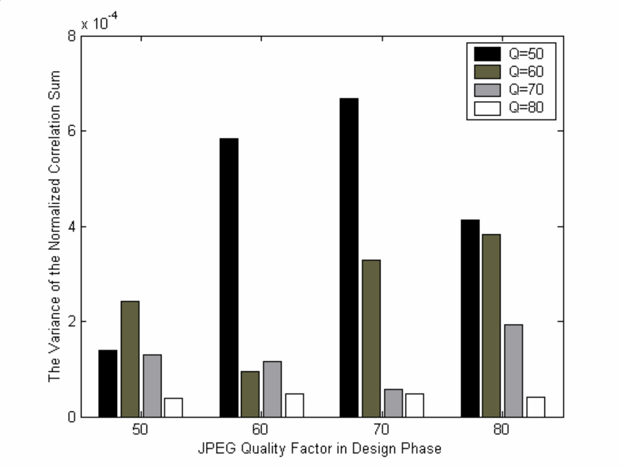
JPEG Quality Factor in Design Phase	Original Scheme			Simplified Scheme	
	No. of Selected Coeff. by Stage 1	No. of Selected Coeff. after Stage 2	Estimated P_{error} after Stage 2	No. of Selected Coeff.	Estimated P_{error}
50	9270	7359	0.0000e+000	7877	3.0992e-246
60	11743	8972	0.0000e+000	10130	0.0000e+000
70	15912	13105	0.0000e+000	13708	0.0000e+000
80	22931	18885	0.0000e+000	20120	0.0000e+000

Table 5.5. The computation complexity (coefficient processing time) for Lena.

JPEG Quality Factor in Design Phase	Number of coefficients processed		
	Original Scheme		Simplified Scheme
	Stage 1	Stage 2	
50	64512	3152520	73629
60	64512	5134207	74108
70	64512	10641870	75139
80	64512	21277960	76366

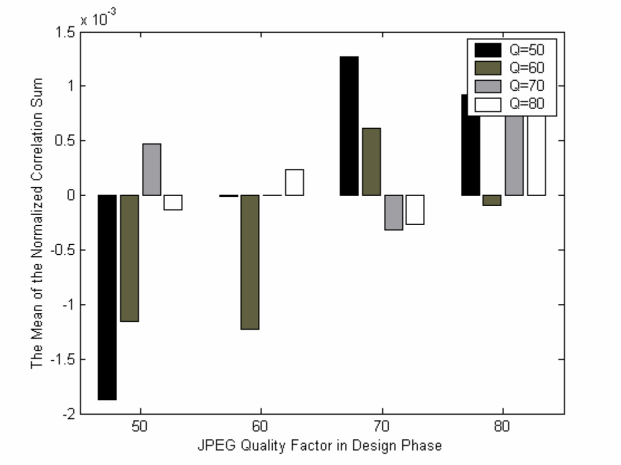


(a)

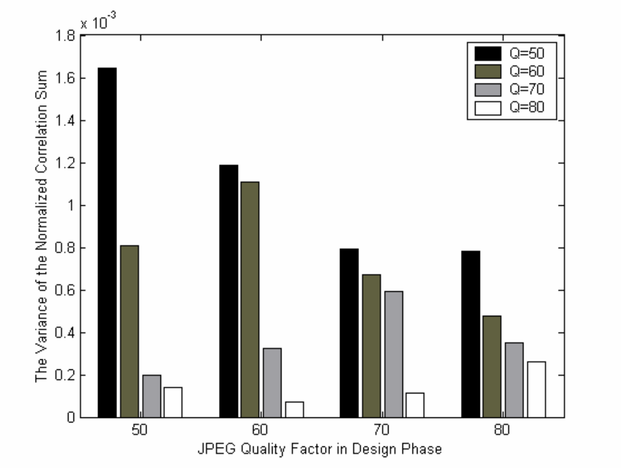


(b)

Fig. 5.3. The mean and variance of the normalized correlation sum after JPEG attacks at different quality factors for watermarked image Lena: (a) Mean $E\{c \mid H_1\}$ and (b) Variance $\text{Var}\{c \mid H_1\}$.

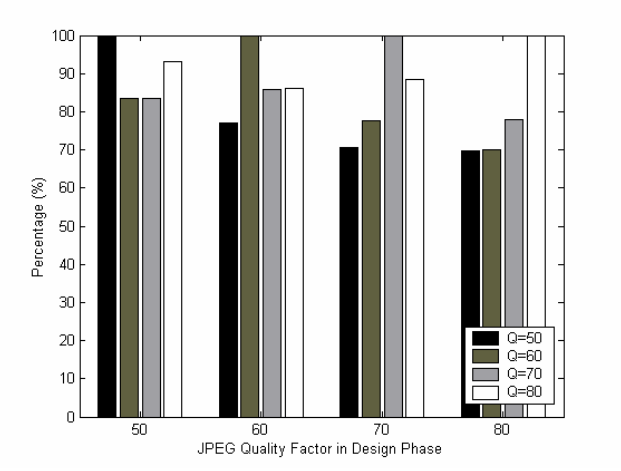


(a)

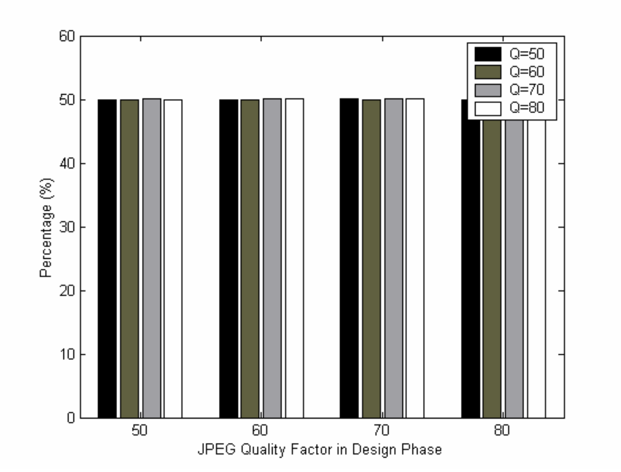


(b)

Fig. 5.4. The mean and variance of the normalized correlation sum after the JPEG attacks (at four different quality factors) for the unwatermarked image Lena: (a) Mean $E\{c \mid H_0\}$ and (b) Variance $\mathrm{Var}\{c \mid H_0\}$.

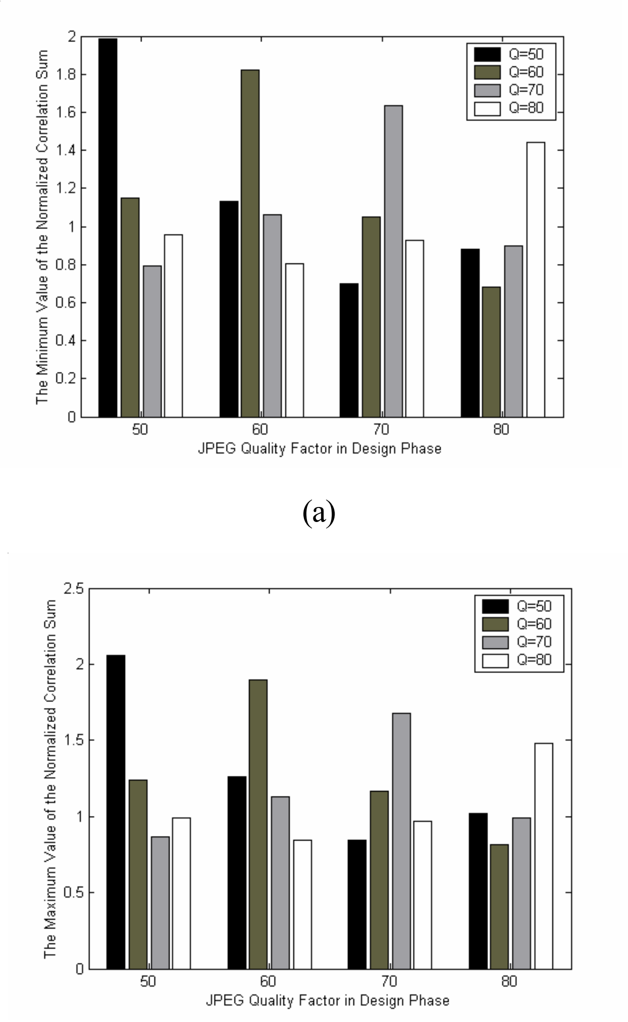


(a)



(b)

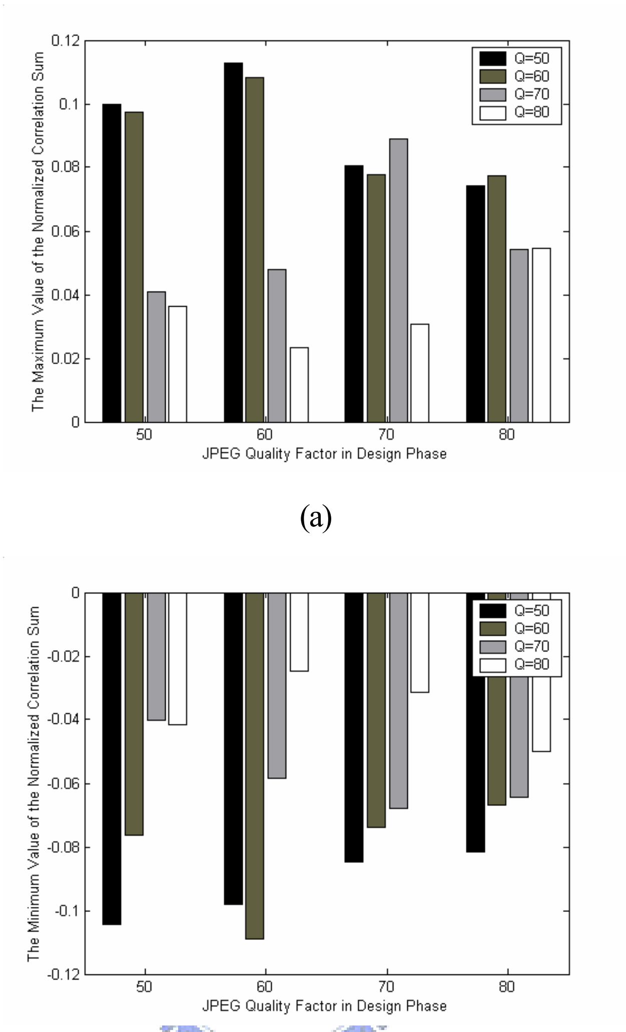
Fig. 5.5. The percentage of correctly decoded coefficients at the detector after JPEG attacks for image Lena: (a) Watermarked and (b) Unwatermarked.



(a)

(b)

Fig. 5.6. The maximum and minimum values of the normalized correlation sum $E\{c \mid H_0\}$ after the JPEG attacks (at four different quality factors) for the watermarked image Lena: (a) Maximum and (b) Minimum.



(a)

(b)

Fig. 5.7. The maximum and minimum values of the normalized correlation sum $E\{c \mid H_0\}$ after the JPEG attacks (at four different quality factors) for the unwatermarked image Lena: (a) Maximum and (b) Minimum.

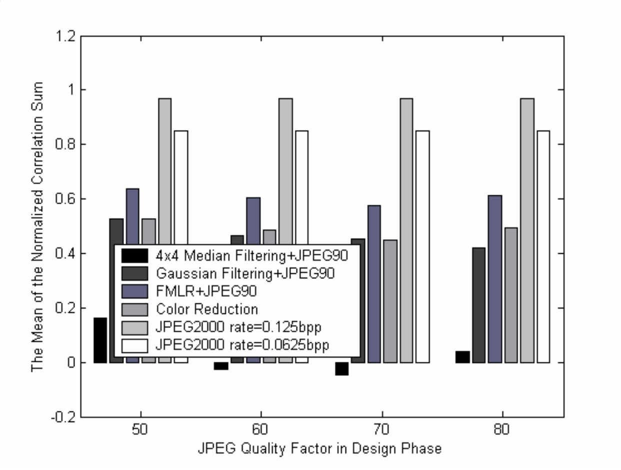


Fig. 5.8. The mean of the normalized correlation sum after various signal processing attacks for watermarked Lena.