Watermarking approach to embedded signature-based authentication by channel statistics

Zhi-Fang Yang
Wen-Hsiang Tsai
National Chiao Tung University
Department of Computer and Information Science
Hsinchu, Taiwan 300
Republic of China
E-mail: gis82504@cis.nctu.edu.tw
whtsai@cis.nctu.edu.tw

Abstract. Embedded signature-based authentication is a highly promising watermarking method for solving high-level authentication problems. This work proposes such a watermarking approach using channel statistics evaluated from a reference watermark. The basic concept is to consider the problem to be an example of communications with side information, which is the reference watermark embedded in the host with the signature at the sender site. At the receiver site, the reference watermark is first extracted to determine channel statistics, which are used to generate information about signature reliability. Three kinds of information are derived, including a reliability measure, an entropy measure, and an authenticated signature. Experimental results establish that the measures of reliability and uncertainty are meaningful, and that the embedded signature can survive high-quality JPEG compression and manipulations such as negation, cropping, replacement, and modification.
[DOI: 10.1117/1.1556394]

Subject terms: watermarking; authentication; embedded signature; channel statistics; reliability; uncertainty.

Paper 020226 received Jun. 5, 2002; revised manuscript received Oct. 9, 2002; accepted for publication Oct. 10, 2002.

1 Introduction

Currently, a mass of multimedia data can be produced and rapidly transmitted using modern digital technologies. Naturally, solutions to digital copyright protection are required urgently to tackle the problem of unauthorized copying and distribution. The research field of copyright marking, concerned with inserting copyright information into host documents, has thus attracted extensive interest in the last decade.[1,2]

In the early 1990s, the methods proposed for copyright marking involved the property of robustness, and robust watermarking algorithms were extensively studied.[1–3] The aim was to make the destruction of embedded watermarks impractical without severely degrading visual quality. Recently, attention has been switched from robust watermarking to fragile watermarking,[3] in which the main aim is to authenticate the host using fragile watermarks.

With respect to watermarking-based authentication, approaches developed so far fall into two main categories, namely fragile watermarking methods and robust watermarking-based methods.[3] Fragile watermarking is a more mature field, since it was studied earlier in the history of visual authentication.[4] Whether or not data is modified, a ''yes-or-no'' question can be answered by fragile watermarking. However, to distinguish different types of distortions, the high-level authentication problem is more desirable for authenticating multimedia data. In Ref. 5, observing modifications of a fragile watermark in different localized spatial and frequency regions make solving the problem possible. But sometimes it is not easy to distinguish new kinds of distortions based on preselected features, especially when only semantic meanings of distortions are described. A similar viewpoint can also be found in Ref. 3, in which robust watermarking is recommended and the watermark is suggested to be a low-resolution image of the host. Generally, the watermark in the second category is a signature of the host embedded in the host at the sender site and can be extracted to authenticate the host at the receiver site.

Research into this topic is still in its infancy.[3] Related work[6–8] has revealed two problems particularly worthy of investigation.[3] One concerns the accuracy of the information used by the signature in summarizing the host. The other concerns the accuracy of the extracted signature at the receiver site. The former is an issue of signature design, and the latter is one of robustness. Generally, more data contained in the signature yield greater accuracy. However, more data hidden in the host yield lower robustness. That is, a tradeoff exists between these two kinds of accuracy. Based on previous work, a compromise involves finding how many data can be reliably embedded in the host; this problem is called the capacity problem.[3] However, the capacity problem remains unsolved[3,9] because of a lack of ability to model arbitrary manipulations that might occur during transmission or other applications.

This work uses a reference watermark to derive channel statistics to address the issue of the accuracy of the embedded signature. The basic concept is to view the problem as one of communications with side information,[14] which is designed to be the reference watermark embedded in the host with the signature. The underlying motivation of this approach is as follows. Comparing accurate original data with received data is the key to successful authentication. However, the extracted signature just cannot be treated as

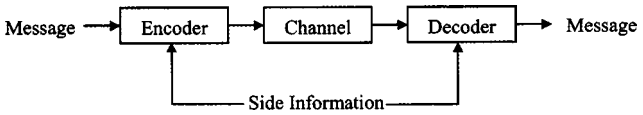


Fig. 1 A communications model with side information.

an accurate one because the capacity problem, namely, how many bits can be reliably embedded, is still open. In this study, the original reference watermark is available at the receiver site. That is, authentication of the reference watermark is guaranteed. The authentication results of the reference watermark are thus used to generate information about signature reliability. In this study, three kinds of information are provided, including a reliability measure, an entropy measure, and an authenticated signature. The former two measures are used to inform users at the receiver site about some quantitative measures regarding the reliability and uncertainty of the transmitted image.

Research on using reference or pilot watermarks to derive channel statistics can also be found in Refs. 10–12. In Voloshynovskiy et al.,[10] a key-dependent reference pilot watermark is used to estimate channel state, the noise distribution, after attack. Their goal is to design an optimal decoder for images under attacks, including geometrical attacks and additive non-Gaussian noise. In Kundur and Hatzinakos,[11] a reference sequence is utilized for channel estimation that is used to adjust the receiver filter to maximize detection reliability of the watermark. In addition, a new watermarking channel, a nonstationary parallel binary symmetric channel (BSC) model, is introduced. In Bäuml, Eggers, and Huber[12] a securely embedded pilot signal is exploited to estimate the transformation of the sampling grid. The channel estimate is then used to invert the desynchronization attack before applying the Scalar Costa Scheme.

The rest of this work is organized as follows. Section 2 formulates the problem to explain how robust watermarking-based authentication can be viewed as communications with side information. The computation of channel statistics is also discussed. Section 3 describes the proposed approach. Section 4 gives some experimental results and related discussion. Conclusions and recommendations for future work are finally made in Sec. 5.

2 Authentication as Communications with Side Information

According to Fig. 1, a communications system encodes a message at the sender site, sends it through a channel, and decodes it at the receiver site. The side information is used by both the encoder and the decoder to improve the transmission rate or enhance the detectability of the signature embedded in the message.

Figure 2 illustrates the proposed approach. The signature

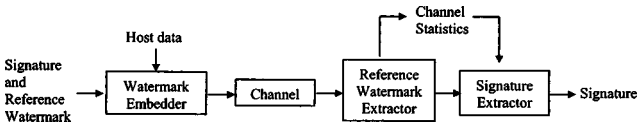


Fig. 2 The model of the proposed approach.

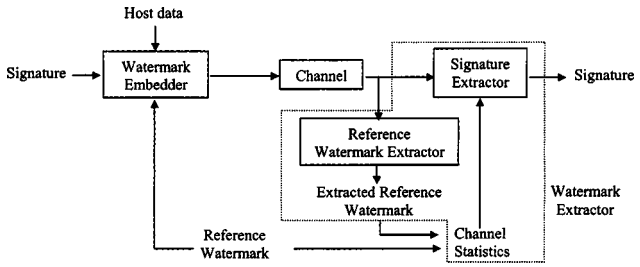


Fig. 3 The proposed approach viewed as communications with side information.

is embedded in the host data with a reference watermark. After transmission through the channel, the embedded reference watermark is extracted and used to determinethe channel statistics. The signature is then detected and used to authenticate the transmitted host data.

In this work, the side information, which is a reference watermark, is known by both the watermark embedder at the sender site and the watermark extractor at the receiver site. The channel statistics are provided to the signature extractor using the side information. This mapping naturally leads to an example of communications models with side information, as shown in Fig. 3.

The concept of the proposed approach is thus clarified: a reference watermark is used as side information to derive channel statistics that are then used to reliably extract the signature to enable authentication. Two assumptions are made without a loss of generality. The first is that manipulations in the channel cannot distinguish the embedded signature and the embedded reference watermark in the host data. The second is that the manipulations on the host are locally consistent. Thus, local channel statistics can be analyzed by both a designed reference watermark at the sender site and the corresponding extracted reference watermark at the receiver site.

3 Proposed Approach

3.1 *Watermark Design and Embedding*

The watermark is composed of a signature and a reference watermark. The total size cannot be so large as to destroy the transparency of the image. Moreover, the number of reference watermark bits cannot be too few with respect to the number of signature bits; otherwise, the computed channel statistics are unreliable in the local area. A binary bit stream is used as the reference watermark. It is used to detect channel statistics based on the two assumptions concerning manipulations stated before. The signature is generated by thresholding the host image according to a threshold preselected manually for each individual host image. That is, a set of threshold candidates is tested for a specific host image. The threshold candidate with the clearest signature is selected according to the subjective judgment of the user. Doing so is an attempt to produce a distinct signature for human visibility. Such a design avoids the difficulty of summarizing information contained in the host data. Other kinds of signatures are also suitable.

After being created, the signature and the reference watermark are randomly permuted using a permutation key to increase security and distribute the reference watermark

bits as uniformly as possible. The permutation includes two stages. In the first stage, the signature is randomly permuted alone. After the permutation, two equally long bit streams of zeros and ones, the reference watermark, are appended to the permuted signature bit stream. Then, the three kinds of bit streams, the permuted signature, the zeros, and the ones, are divided into equal numbers of substreams. Substreams with the same order of corresponding bit streams are treated as a subgroup. In the second stage, all the subgroups are permuted, respectively. The final complete watermark is created when all permutations are performed.

The embedding method is basically the quantization-based method developed by Kundur and Hatzinakos[11]: the host image is transformed using a wavelet transform, and the watermark is embedded bit by bit in the randomly chosen detail coefficients. Two modifications are made to improve robustness: one is not to embed watermark bits at the highest resolution; the other is to set the quantization parameter as a constant. This study suggests that if many wavelet levels are applied, the constant quantization parameter at a middle resolution, as computed in Ref. 5, is selected. However, if fewer levels are applied, the selected constant quantization parameter is the one calculated in Ref. 5 at the lowest resolution. Experiments performed for this study reveal that if the number of levels is five, the selected quantization parameter should be 2^3; if the number of levels is three, the selected quantization parameter should also be 2^3. Briefly, the embedding method is designed to be insensitive to smaller modifications and to prevent the host from severe visual distortion.

The following rules summarize the proposed embedding method. Notation, statements, and rules are consistent with those developed in Ref. 5. For each watermark bit $w(i)$, a wavelet coefficient $f_{k,l}(m,n)$ is randomly selected, where $k=h$, v, or d, which stand for horizontal, vertical, and diagonal detail coefficients, respectively; $l=1,2,...,L$ and is a specific resolution, and (m,n) specifies a spatial location. Notably, each wavelet coefficient can be selected only once to prevent the production of errors by repeatedly embedding one coefficient. The embedding rules are as follows:

$$f_{k,l}(m,n) := \begin{cases} f_{k,l}(m,n)+\Delta & \text{if } f_{k,l}(m,n)\leqslant 0 \text{ and } Q[f_{k,l}(m,n)]\neq w(i) \\ f_{k,l}(m,n)-\Delta & \text{if } f_{k,l}(m,n)\geqslant 0 \text{ and } Q[f_{k,l}(m,n)]\neq w(i), \\ f_{k,l}(m,n) & \text{if } Q[f_{k,l}(m,n)]=w(i) \end{cases} \tag{1}$$

where $:=$ is the assignment operator; Δ represents the unique quantization parameter discussed in this section; and $Q(.)$ is a quantization function which is defined as follows:

$$Q[f_{k,l}(m,n)]=\begin{cases} 0 & \text{if } \left\lfloor \dfrac{f_{k,l}(m,n)}{\Delta} \right\rfloor \text{ is an even number} \\ 1 & \text{otherwise} \end{cases}. \tag{2}$$

3.2 *Watermark Extraction and Authentication*

At the receiver site, localizing the watermark is rather simple by the coefficient selection and permutation keys. The transmitted host image is first transformed into the predefined wavelet domain. Then the coefficient selection key is used to locate the wavelet coefficient positions in which the watermark bits are embedded. Later, the permutation key is used to separate the embedded signature bits from the embedded reference watermark bits.

The channel statistics are determined by analyzing the reference watermark bits in a predefined neighborhood of each signature bit. In this study, the neighborhood is an $N\times N$ square centered at the point in which the signature bit is embedded in the host image. The size of the neighborhood should contain sufficient reference watermark bits to make the statistics meaningful. Furthermore, the neighborhood should not be so large that it violates the local statistics of the channel condition. Therefore, the size of the neighborhood should be chosen appropriately. In this work, experiments show that a 21×21 neighborhood is suited to analyze a 512×512 host image. Conditional probabilities $p_{i|j}$ (where i is the value of the extracted reference watermark bit and j is the value of the corresponding original reference watermark bit) are evaluated according to the following equation after the reference watermark bits are located in the neighborhood ofa specific signature bit b:

$$p_{i|j}=\frac{\Sigma_{1\leqslant p,q\leqslant N}(1-e_{p,q}\oplus i)(1-r_{p,q}\oplus j)}{\Sigma_{1\leqslant p,q\leqslant N}(1-r_{p,q}\oplus j)} \quad \text{for } i,j\in\{0,1\}, \tag{3}$$

where $\oplus$ is the XOR operator; $e_{p,q}$ represents an extracted reference watermark bit at position (p,q) in the $N\times N$ neighborhood, and $r_{p,q}$ represents the original reference watermark bit embedded at the same position at the sender site. Four kinds of conditional probabilities, $p_{0|0}$, $p_{0|1}$, $p_{1|0}$, and $p_{1|1}$, are computed, and correspond to pairs of the extracted and original bits (0, 0), (0, 1), (1, 0), and (1, 1), respectively. Notably, Eq. (3) can be easily extended to the case of more signature bits being considered simultaneously.

Since the *a priori* probabilities P_j of the classes C_j, which denote the classes of the signature bits with the value $j\in\{0,1\}$, are unknown at the receiver site in this study, they are set to be identical, without a loss of generality. That is, P_0 and P_1 are both 0.5. Then, the estimated bit value b' of the extracted signature bit b is determined using the Bayes theorem[13] with an additional constraint:

$$b'=\begin{cases} \arg\max\limits_{j\in\{0,1\}} \dfrac{p_{b|j}P_j}{\Sigma_{0\leqslant k\leqslant 1}p_{b|k}P_k} & \text{for } \max\limits_{j\in\{0,1\}} \dfrac{p_{b|j}P_j}{\Sigma_{0\leqslant k\leqslant 1}p_{b|k}P_k}>T \\ \text{unreliable} & \text{otherwise} \end{cases}, \tag{4}$$

where T is the threshold of the additional constraint:

$$\frac{p_{b|j}P_j}{\Sigma_{0\leqslant k\leqslant 1}p_{b|k}P_k}>T.$$

This constraint makes the maximum *a posteriori* probability much larger than the other *a posteriori* probability. The extracted signature bit b' is classified as ''reliable'' if one of the *a posteriori* probabilities passes the test; otherwise, b' is classified as ''unreliable.'' The concept is that the computed value may not be sufficiently representative if the difference between the two *a posteriori* probabilities is insignificant. In this work, T is suggested to be 0.8, following the experimental results. Now, every bit of the signature can be estimated from the channel statistics. A reliability measure, denoted as R, of the extracted signature is then defined as the following statistical value:

$$R=\frac{\Sigma_{1\leqslant i\leqslant L_s}(1-b_i'\oplus b_i')}{L_s}$$

for b_i' is in the ''reliable'' class, (5)

where L_s is the length of the signature stream. In Eq. (5), the number of the extracted signature bits classified as ''reliable'' is divided by the signature length L_s. This percentage measure R is then used to represent the reliability of the extracted signature. The uncertainty of the corresponding channel statistics of the extracted signature is defined as the entropy H. The entropy H of an extracted signature is determined as follows:

$$H=\frac{\Sigma_{1\leqslant i\leqslant L_s}\Sigma_{0\leqslant i,j\leqslant 1}(-p_{i,j}\log_2 p_{i,j})}{L_s}, \quad (6)$$

where $p_{i,j}$ are the joint probabilities determined from the following equation:

$$p_{i,j}=\frac{\Sigma_{1\leqslant p,q\leqslant N}(1-e_{p,q}\oplus i)(1-r_{p,q}\oplus j)}{\Sigma_{1\leqslant p,q\leqslant N}[(1-r_{p,q}\oplus 0)+(1-r_{p,q}\oplus 1)]}$$

for $i,j\in\{0,1\}$, (7)

where $e_{p,q}$ and $r_{p,q}$ represent an extracted reference watermark bit and the corresponding correct bit at position (p,q) in the tested $N\times N$ neighborhood, respectively. This entropy measure H reflects the uncertainty of the channel in the local area. Finally, the watermarked image is compared to the extracted signature by visual judgments, by typical matching techniques, or by ad hoc approaches, to classify various manipulations. In this work, for simplicity, visual judgments are adopted in experiments about authentication and classification of various manipulations.

4 Experimental Results and Discussion

The experiments involve evaluating the reliability of the extracted signature and demonstrating the authentication effectiveness of the proposed approach. First, we evaluate the reliability measure R, computed from Eq. (5). More specifically, the most interesting question is, ''Is the measure of reliability based on channel statistics R itself reliable?'' Two experiments were conducted to answer the question. One compared the reliability measure R to the hit ratio of the corresponding extracted signature bits. The hit ratio A of an extracted signature is determined as follows:

$$A=\frac{\Sigma_{1\leqslant i\leqslant L_s}(1-b_i'\oplus b_i^0)}{\Sigma_{1\leqslant i\leqslant L_s}b_i'}$$

for b_i' is classified into the ''reliable'' class, (8)

where b_i' is a reliable bit and the i'th bit of the extracted signature, and b_i^0 is the i'th bit of the original signature stream. The hit ratio A is the percentage of the reliable signature bits that arecorrect. The signature bits detected as reliable can be confidently held to contain a large percentage of correct bits if a reliability measure R appears with a large hit ratio. That is, the reliable signature bits can be trusted to perform authentication under such circumstances.

The other experiment concerning the evaluation of reliability compared the reliability measure R to the uncertainty of the corresponding channel statistics H. As is well known, entropy represents uncertainty in information; larger entropy values imply greater uncertainty of information. Accordingly, when a reliability measure R appears with a low entropy value, little uncertainty surrounds the detected signature bits that are recognized as reliable. That is, the values of the reliable signature bits can be believed.

The experiments concerning authentication were performed with the aim of demonstrating some authentication abilities of the proposed method. In this work, four kinds of manipulations are considered—negation, cropping, replacement, and modification. The desired aim is to extract correct signatures from transmitted images. Therefore, the correct signatures can be used to authenticate visually whether the transmitted images are attacked by the four manipulations. The following paragraph introduces parameters used in the experiments.

The experiments were performed on monochrome images with a size of 512×512. Three images were used—Airplane, Lena, and Baboon—and were selected to represent three kinds of images: those containing large smooth areas, those containing both smooth and detailed areas, and those containing huge amounts of details. The signatures are generated by thresholding the monochrome images to black and white ones and then reducing them to a size of 128×128 by selecting the first pixel of every four pixels in a row from top to bottom, left to right. In the experiments, a gray-level threshold of 128 was used for the images Lena and Baboon. A gray-level threshold of 185 was selected for the image Airplane. The Haar wavelet transform is used and the resolution levels are chosen as three. The length of the reference watermark was set to a quarter of the signature size in these experiments. After the permutations specified in Sec. 3.1 are applied, the watermark is embedded into the transformed images using a unique quantization parameter, 2^3, according to the method presented in Ref. 5. The PSNRs, which are calculated from the formula given in Ref. 5, of the embedded images in the experiments are all between 40 and 50. At the receiver site, the watermark is extracted from the transmitted images using the permutation keys, following what is described in Sec. 3.2. And the channel statistics are analyzed in a 21×21 neighborhood around each signature bit position. The threshold T is set at 0.8 in the experiments. Then, the value of each signature bit can be computed from the corresponding channel statistics.

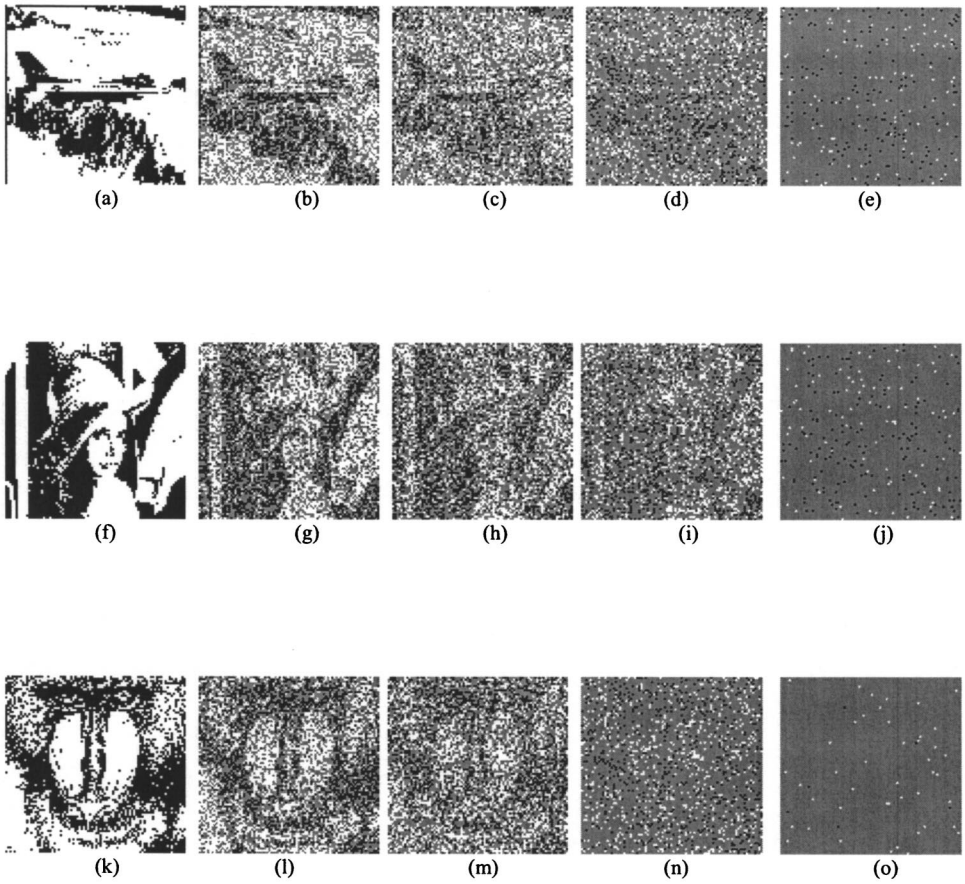


Fig. 4 Examples of signatures extracted from JPEG images used in experiments on hit ratios and entropies. The reliable extracted signature bits are represented by black and white pixels, which indicate the bit values 0 and 1, respectively. The unreliable signature bits are represented by gray pixels. Three images, Airplane, Lena, and Baboon, are shown for different reliability measures R and entropy measures H. For the Airplane image, (a) original; (b) $R=0.52$ and $H=1.24$; (c) $R=0.43$ and $H=1.72$; (d) $R=0.24$ and $H=1.78$; and (e) $R=0.03$ and $H=1.87$. For the Lena image, (f) original; (g) $R=0.54$ and $H=1.44$; (h) $R=0.5$ and $H=1.69$; (i) $R=0.33$ and $H=1.75$; and (j) $R=0.03$ and $H=1.86$. For the Baboon image, (k) original; (l) $R=0.55$ and $H=1.27$; (m) $R=0.5$ and $H=1.67$; (n) $R=0.18$ and $H=1.81$; and (o) $R=0.02$ and $H=1.92$.

Finally, the computed signatures are visually compared with the transmitted images to perform the authentication.

With respect to the properties of the reliability measure R, a series of JPEG operations with various compression ratios are used. These JPEG images are generated by adjusting parameters pertaining to functions of saving images provided by the software, MATLAB. In the experiments, the images are compressed with lost data percentages between 3 and 98%, which is the ratio of the size difference between an original image and its corresponding JPEG image to the size of the original image. These images exhibit various degrees of distortion, as therefore do the embedded signatures. Figure 4 shows some examples of extracted signatures from these JPEG images. In the figure, the reliable signature bits are represented by black and white pixels, which represent the bit value zero and the bit value one, respectively. The unreliable signature bits are represented by gray pixels.

Figure 5 compares the measures of reliability with hit ratios. Larger reliability measures are associated with larger hit ratios for all three images. Therefore, such a reliability measure based on channel statistics is meaningful. In particular, the hit ratios exceed 90% when the reliability exceeds 0.5, and under such circumstances, the lost data percentages can be 60% for JPEG images of Airplane and Lena and 38% for the Baboon image. The reliability measures approach zero when the lost data percentages increase to a high value. That is, the distortion caused by JPEG compression so severely damaged the images that very few or no extracted signature bits can be considered to be reliable. Hit ratios also reflect the same situation: the hit ratios are either smaller than 0.5 or are 0 under such circumstances. Notably, the values of the reliable signature bits are random if the corresponding hit ratio is one half. The numbers of zeros in the hit ratios of the three images are quite different. The largest number of zeros appears in the Baboon image. This result is because the image Baboon includes many details, which are more easily distorted by JPEG compression, than those in the other two images. Under such circumstances, channel statistics tend to have equal shares and then the extracted signature bits are judged as unreliable. Therefore, the corresponding hit ratio

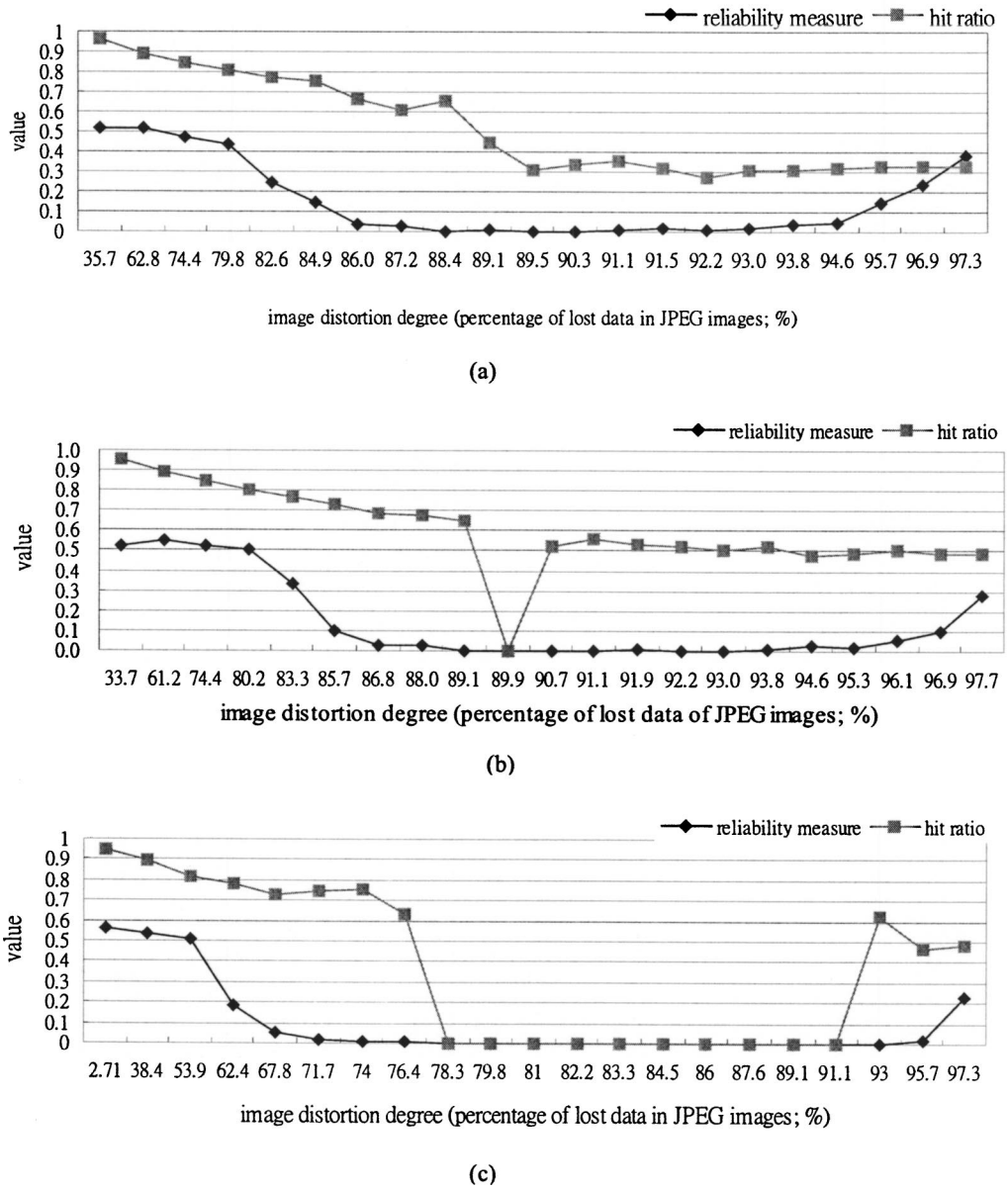


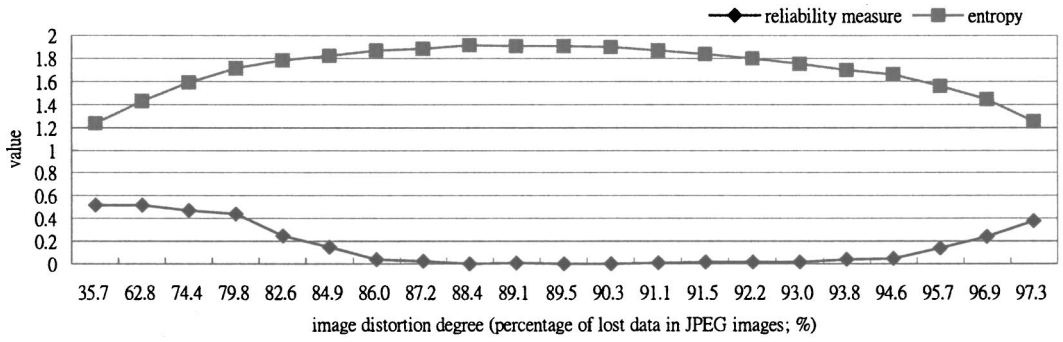
Fig. 5 Experimental comparisons between reliability measure R and hit ratio A obtained by testing a series of JPEG images: (a) Airplane; (b) Lena; and (c) Baboon.

and its reliability measure are both zero. With respect to the image of Lena, which has fewer details than the Baboon image, the hit ratio has fewer zeros than the Baboon image. The image Airplane, which is smooth in most areas, contained no zero hit ratios in the experiments.

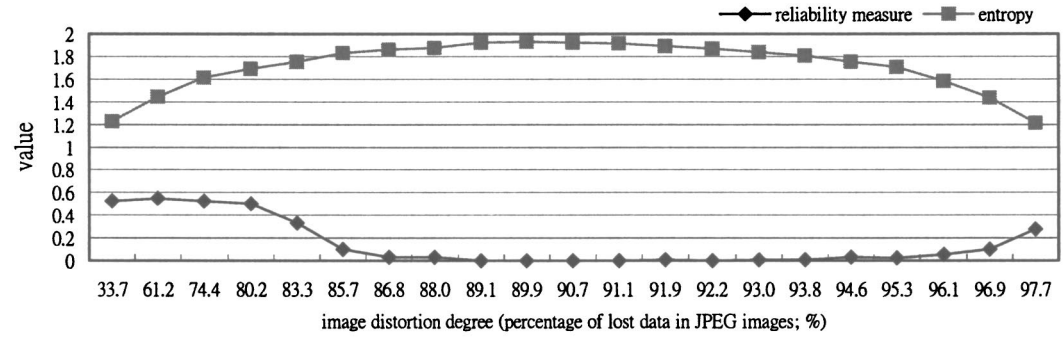
Under sever image distortion, the values of the extracted reference bits and the signature bits are unpredictable (we may consider in this case that $p_{0|0} \sim p_{0|1} \sim p_{1|0} \sim p_{1|1}$) and the conclusion about the reliability of these bits will be in most of the cases unreliable. Then users will know that this transmission is not unreliable. That is, the unreliable information is also a kind of information. Note that such information cannot be provided by using an embedded signature alone.

Figure 6 compares the measures of reliability with the entropy of the signatures. Notably, larger reliability measures accompany lower entropy in the experimental results; that is, the informational uncertainty is quite low for those computed reliable signature bits. When the reliability measures are near zero, the corresponding entropy values are rather high (the maximum entropy value is 2 in the experiments), which implies that the uncertainty of the information is high. Therefore, the channel statistics-based reliability measure is meaningful. Moreover, the curve shapes of the entropy values are rather symmetrical; that is, entropy is low at both smaller and larger degrees of distortion. Therefore, the channel statistics reveal high certainty in these two situations.

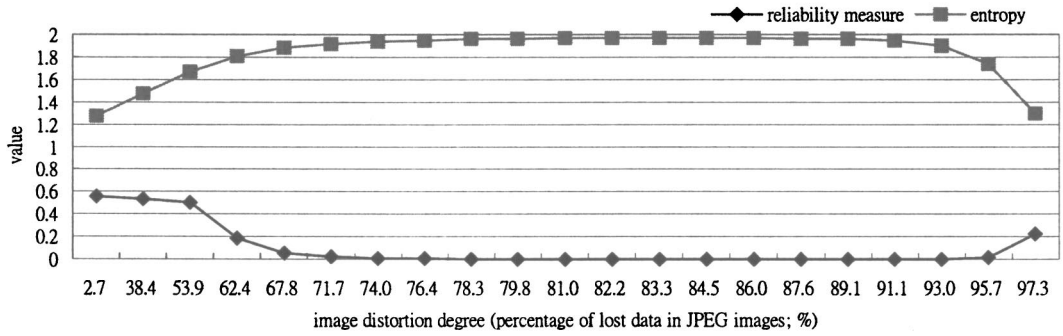
Figures 7–12 show experimental results concerning authentication. For instance, Fig. 7(a) is the original host image. After the watermark is embedded, the host image is manipulated by the negation operation. As shown in Fig. 7(b), the airplane seems to be imaged at night. Now, the power of the approach based on channel statistics can be



(a)



(b)



(c)

Fig. 6 Experimental comparisons between reliability measure *R* and entropy *H*, obtained by testing a series of JPEG images: (a) Airplane; (b) Lena; and (c) Baboon.

illustrated as follows. In Fig. 7(c), the extracted signature is obtained by directly retrieving the bit values stored in the signature positions, and the airplane still looks as if it is flying at night. However, after the proposed channel-statistics-based approach is used to detect the embedded signature, the result correctly reveals the airplane actually flies by day. Figures 8 and 9 show similar situations. All three images give successful results.

The proposed approach should now be compared to the conventional approach of embedding the signature repeatedly without any side information such as channel statistics. Two points are worthy of note. The first arises from experiments on negation attack. The traditional repeating method clearly cannot survive such an attack, since no knowledge is available concerning what happens in the transmission. Sometimes, a conventional voting mechanism can be used to guess the transmission conditions. However, voting-based methods fail if the negation manipulation is performed. Under such circumstances, the approach based on channel statistics is superior, because the channel statis-

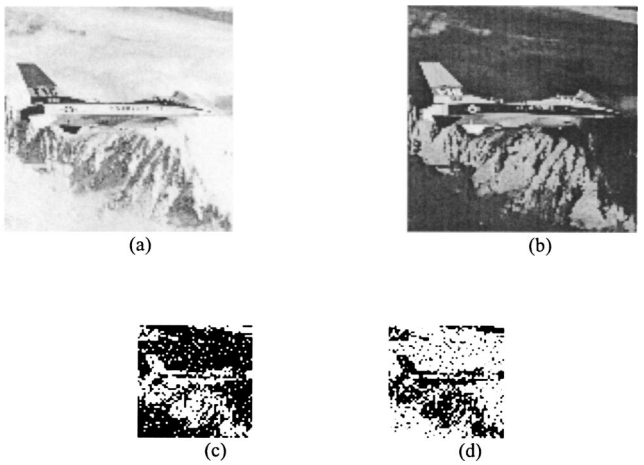


Fig. 7 Example of negation attack: (a) the original host image Airplane; (b) negation of the watermarked image; (c) signature directly extracted without applying the proposed mechanism; and (d) signature extracted by the proposed approach.

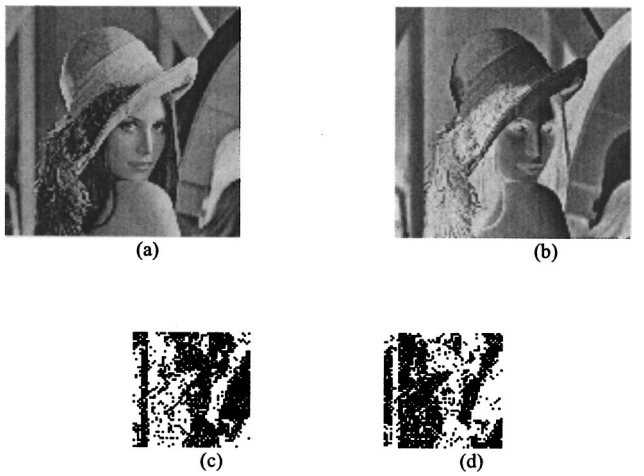


Fig. 8 Example of negation attack: (a) the original host image Lena; (b) negation of the watermarked image; (c) signature directly extracted without applying the proposed mechanism; and (d) signature extracted by the proposed approach.

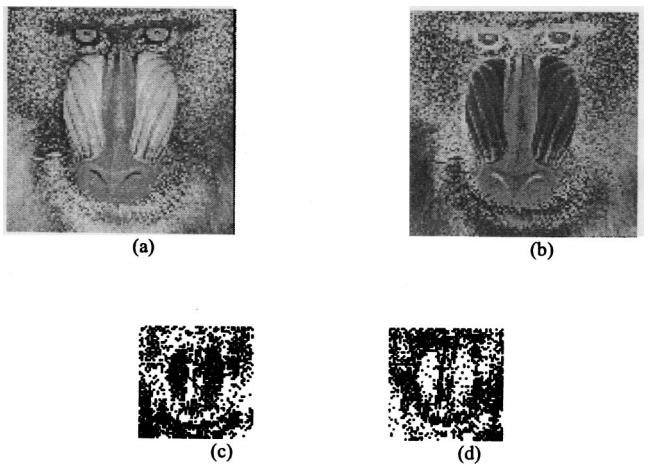


Fig. 9 Example of negation attack: (a) the original host image Baboon; (b) negation of the watermarked image; (c) signature directly extracted without applying the proposed mechanism; and (d) signature extracted by the proposed approach.

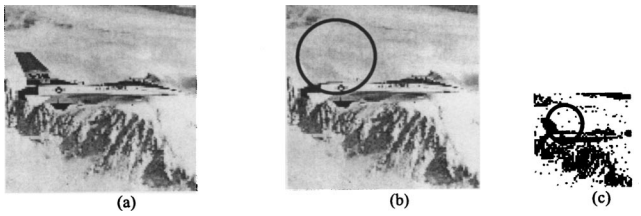


Fig. 10 Example of cropping attack: (a) original host image; (b) attacked image (the top rear wing is cropped from the watermarked image); and (c) extracted signature showing the top rear wing.

tics can be used to collect information about transmission quality. The second point is that the repeating method uses more space than the proposed approach. As described in this section, the reference watermark size is only one quarter of the signature size. However, the conventional repeating method uses double the signature size even if the signature is embedded only twice.

Other kinds of attacks are also examined, including cropping, replacement, and modification. In Fig. 10, the top wing on the rear of the airplane is cropped after the watermark is embedded. Figure 10(b) shows the cropped image. However, the extracted signature presents the original picture of the airplane, which is different from the transmitted one. Thus, the authentication can be performed successfully. In Fig. 11(b), the original white hat is replaced by a black one. The replacement manipulation can be detected by comparing the attacked image with the black hat to and the extracted signature with a white hat. The areas near the nose of the baboon in Fig. 12(a) are modified into interlaced bands of black and white, as shown in Fig. 12(b). By comparing the attacked image in Fig. 12(b) with the extracted signature in Fig. 12(c), the original baboon is shown to differ from the transmitted image in Fig. 12(c). Therefore, the authentication has also been performed successfully.

5 Conclusions

Authentication based on embedded signatures constitutes a growing field of research. In previous work, robust watermarking-based methods were designed to tackle this problem by focusing on the robustness of embedding signatures and generating efficient signatures. This study seeks to develop an approach based on channel statistics, which gives useful information for authentication, including a measure of reliability of the extracted signature and a mea-

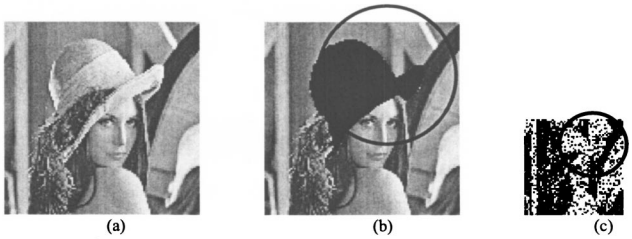


Fig. 11 Example of replacement attack: (a) original host image; (b) attacked image (the hat is replaced by a black one in the watermarked image); and (c) extracted signature showing the original hat color as white.

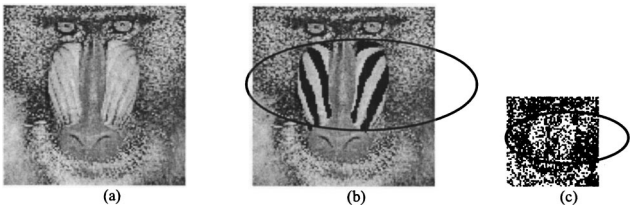


Fig. 12 Example of modification attack: (a) original host image; (b) attacked image (the face of the baboon is modified in the watermarked image); and (c) extracted signature showing the original baboon face.

sure of the uncertainty of the channel status. The goal is to extract reliable information based on the transmission quality at the receiver site rather than to consider the problem of robustness at the sender site. The watermarking-based authentication problem is thus viewed as an example of communications with side information. The important point is to use a reference watermark as side information to derive channel statistics.

The experimental results concerning hit ratios and uncertainties have shown that such a method based on channel statistics can be used to describe reasonably the channel status. The experimental results reveal that the proposed approach can be used to authenticate transmitted images manipulated by high-quality JPEG compression, negation, cropping, replacement, or modification. Both the comparison of the watermarked image with the extracted signature and the classification of various manipulations were achieved by visual judgments. Future research is suggested to develop more objective methods, such as methods based on typical matching techniques, to do the comparison and the classification. Other future research may be directed to developing more robust methods of embedding, deriving the optimal size of the neighborhood used to authenticate the signature, and generating more representative signatures.

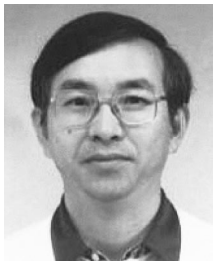
Acknowledgment

This work was supported by the MOE Program of Excellence under Grant No. 89-E-FA04-1-4.

References

1. F. Hartung and M. Kutter, ''Multimedia watermarking techniques,'' *Proc. IEEE* **87**(7), 1079–1107 (July 1999).
2. F. A. P. Petitcolas, R. J. Anderson, and M. G. Kuhn, ''Information hiding—a survey,'' *Proc. IEEE* **87**(7), 1062–1078 (July 1999).
3. F. Bartolini, A. Tefas, M. Barni, and I. Pitas, ''Image authentication techniques for surveillance applications,'' *Proc. IEEE* **89**(10), 1403–1418 (Oct. 2001).
4. G. L. Friedman, ''The trustworthy digital camera: restoring credibility to the photographic image,'' *IEEE Trans. Consumer Electron.* **39**, 905–910 (Nov. 1993).
5. D. Kundur and D. Hatzinakos, ''Digital watermarking for telltale tamper proofing and authentication,'' *Proc. IEEE* **87**(7), 1167–1180 (July 1999).
6. J. Fridrich, ''Image watermarking for tamper detection,'' *Proc. ICIP98, IEEE Int. Conf. Image Process.* **2**, 404–408 (1998).
7. C. Y. Lin and S. F. Chang, ''Issues and solutions for authenticating MPEG video,'' *Proc. SPIE* **3657**, 54–65 (1999).
8. M. Wu and B. Liu, ''Watermarking for mage authentication,'' *Proc. ICIP98, IEEE Int. Conf. Image Process.* pp. 437–441 (1998).
9. P. Moulin, ''The role of information theory in watermarking and its application to image,'' *Signal Process.* **81**(6), 1121–1139 (June 2001).
10. S. Voloshynovskiy, F. Deguillaume, S. Pereira, and T. Pun, ''Optimal adaptive diversity watermarking with channel state estimation,'' *Proc. SPIE* **4314**, 673–685 (2001).
11. D. Kundur and D. Hatzinakos, ''Diversity and attack characterization for improved robust watermarking,'' *Signal Process.* **49**(10), 2383–2396 (Oct. 2001).
12. R. Bäuml, J. J. Eggers, and J. Huber, ''A channel model for desynchronization attacks on watermarks,'' *Proc. SPIE* **4675**, 281–292 (2002).
13. K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed., Academic Press, Boston (1990).
14. I. J. Cox, M. L. Miller, and A. L. McKellips, ''Watermarking as communications with side information,'' *Proc. IEEE* **87**(7), 1127–1141 (July 1999).

Zhi-Fang Yang received her PhD degree in computer and information science from Chiao Tung University, Hsinchu, Taiwan in 1999. She was a postdoctoral fellow at the Institute of Information Science, Academia Sinica, Taiwan from December 1999 to December 2000. Since January 2001, she has been a contracted assistant professor in the Department of Computer and Information Science at National Chiao Tung University. Her current research interests include multimedia security and communication, multimedia information processing, and watermarking theory.



Wen-Hsiang Tsai received his PhD degree in electrical engineering from Purdue University in 1979. Dr. Tsai joined the faculty of National Chiao Tung University, Hsinchu, Taiwan in November 1979. He has also been the Head of the Computer and Information Science Department, the Associate Director of the Microelectronics and Information System Research Center, the Dean of General Affairs, and the Dean of Academic Affairs there. He is currently the Vice President of the University. He has been the editor of several academic journals, including *Computer Quarterly* (now *Journal of Computers*), *Proceedings of the National Science Council, International Journal of Pattern Recognition*, and the *Journal of Chinese Engineers*; and the Editor-in-Chief of the *Journal of Information Science and Engineering*. His major research interests include image processing, pattern recognition, computer vision, document image analysis, and information hiding. He has published more than 268 academic papers, including 109 journal papers. He has received many awards from the National Science Council and from the Chinese Image Processing and Pattern Recognition Society. He is a senior member of IEEE and a member and the Director of the Chinese Image Processing and Pattern Recognition Society. He is also a member of the Medical Engineering Society of the Republic of China, the Information Processing Society, and the International Chinese Computer Society.