

一种有意义水印盲提取算法

钟桦 刘芳 焦李成

(西安电子科技大学雷达信号处理国家重点实验室, 西安 710071)

摘要 提出了一种基于匹配滤波的有意义水印盲提取算法. 通过构造基于特征子空间的正交水印信号集, 水印的提取不需要原始图象, 而且误码率可以控制在较低的水平. 基于特征子空间分解特性, 具有简单有效的特点, 同时由于水印信号集由原始图象唯一生成, 并具有一定的不可逆性, 因此可以防止水印攻击者利用非法产生的水印集来伪造有意义的水印. 实验结果证明了该算法具有较好的稳健性和不可见性. 为了对提取字符的可靠性进行评价, 还提出了一种置信度门限. 利用置信度门限可以有效地评估单个字符及整体水印的可靠性, 弥补了一些算法只提取有意义水印, 而不加评估的缺点.

关键词 有意义水印 盲提取算法 匹配滤波 置信度门限

中图法分类号: TP391.41 TP309.7 **文章标识码:** A **文章编号:** 1006-8961(2002)10-1000-05

A Blind Meaningful Watermarking Algorithm

ZHONG Hua, LIU Fang, JIAO Li-cheng

(National Key Lab. for Radar Signal Processing, Xidian University, Xi'an 710071)

Abstract In this paper, a blind meaningful watermarking algorithm is proposed. The host image is first transformed into wavelet domain where the feature set is extracted. Through subspace decomposition of the feature set, the watermark signal sets are generated which holds orthogonality between each two watermark signals and between the watermark signals and the feature set. For the orthogonality between the watermark signals and the feature set, the meaningful watermark can be extracted without resorting to the host image. And for the orthogonality between each two watermark signals, the characters embedded can be retrieved robustly using matched filtering. Because the watermark signal sets are image content dependent, the algorithm security can be improved. In order to measure the reliability of the extracted characters, the concept of credibility threshold is proposed by which the wrong characters can be correctly pointed out. Therefore the credibility of the whole message can also be evaluated, which is another contribution of this paper. In the experiment, we test the algorithm performance under JPEG compression and Additive White Gaussian Noise (AWGN). The results show that the robustness and imperceptibility of the watermark are satisfactory. The credibility threshold also proves its effectiveness.

Keywords Meaningful watermarking, Blind extracting algorithm, Matched filtering, Credibility threshold

0 引言

数字技术的飞快发展使得对电子数据的版权保护技术的需求日益迫切. 作为解决这一需求的一种很有潜力的技术, 近几年在数字水印这一领域的研究非常活跃^[1]. 数字水印是一个把相关信息或签名嵌入媒

体数据的过程, 其中嵌入信息可以是无意义的伪随机噪声^[2~5], 也可以是有意义的标识符, 公司的二维商标, 印鉴等^[6~9]. 显然有意义的水印将具有更大的应用前景. 通过从加水印图象中提取有意义水印信息将达到伪随机噪声水印不可比拟的说明效果.

从稳健性的角度而言, 伪随机噪声水印具有最大的稳健性, 其只相当于嵌入了 1bit 的信息. 随着

基金项目: 国家自然科学基金项目(60073053)

收稿日期: 2001-08-20; 改回日期: 2002-03-04

嵌入信息量的增多,水印的稳健性将会随之下降。有意义水印由多个字符或比特组成,因此其稳健性不可能与伪随机噪声水印相比,但是有意义水印的稳健性却是相当重要的,特别是当以标识符作为水印时,一个误比特就会产生标识错误。因此如何在水印的不可见性、稳健性和嵌入信息量之间采取折衷是设计有意义水印算法的关键问题。

嵌入有意义水印是数字水印技术发展的一个重要趋势。目前提出的很多水印算法都属于这一类。DCT 域水印算法^[6],通过改变系数之间的关系把二值图象或印鉴逐比特地嵌入到主图象中。在这种水印算法中,可以通过重复嵌入水印比特,以达到一定的稳健性。Servetto 等探讨了图象的水印容量问题^[7],并提出用两个随机序列分别代表比特 0 和 1,从而嵌入一个二值标识符序列,检测时则利用匹配滤波来提取嵌入的比特。Marvel 等^[8]在嵌入标识符序列之前,先对其进行纠错编码,从而有效地提高了嵌入序列的稳健性。黄继武等指出匹配滤波是较多重嵌入和纠错编码更为有效的方法^[9]。

基于特征子空间的水印信号集产生方法,利用水印信号与原始图象信号之间天然的正交性有效地实现了有意义水印的盲提取,同时由于水印信号之间的正交性,水印提取的误码率较低。因为水印信号集由原始图象唯一生成,并具有一定的不可逆性,从而可以防止水印攻击者利用非法产生的水印集来伪造有意义的数字水印^[4,5]。大多数算法^[7~9]并没有对所提取水印的可靠性进行分析。为此提出一种置信度门限,可以精确地对每一个字符的可靠性进行度量,并进一步判断整体水印的置信度。实验结果证明了其有效性。

1 水印信号集的产生

水印信号集由多个序列组成,每一个序列,即入口,代表一个特定的字符(ASCII 码)。由于一个字符由 6 个比特来唯一表示,从这一点来看,水印信号集可以传输更多的内容。为实现字符的准确提取,要求水印信号必须是最大可分离信号。显然满足正交性的水印信号集是最佳的选择。基于特征子空间的水印产生算法能够自然地满足这一要求。

设原始图象的特征集为 $\{I_i\}$, $\{I_i\}$ 是变换域系数的子集^[3,4]。为确保水印的安全性, $\{I_i\}$ 必须经过扰乱处理^[6],扰乱处理使用密钥 Key。把特征集 $\{I_i\}$ 分成 M 个维数为 n 的矢量,用 I_j 表示, $j=1, 2, \dots, M$ 。

对 I_j 的协方差矩阵 R_j 进行特征分解^[10]

$$R_j = P_j \Lambda_j P_j^H \quad (1)$$

其中, P_j 是由归一化特征矢量所构成的正交矩阵,“H”表示共轭转置。 $\Lambda_j = \text{diag}(\lambda_j(1), \lambda_j(2), \dots, \lambda_j(n))$ 是特征值矩阵。对于图象信号,可以假设 $\lambda_j(1) > \dots > \lambda_j(K) > \lambda_j(K+1) = \dots = \lambda_j(n) = 0$, 因此 I_j 可以写成

$$I_j = \sum_{k=1}^K \sqrt{\lambda_j(k)} P_j(k) \quad (2)$$

其中, $P_j(k)$, $k=1, \dots, K$ 称为信号特征矢量,记作 P_j^s , 而 $P_j(k)$, $k=K+1, \dots, n$ 称为噪声特征矢量,记作 P_j^N 。根据正交矩阵的特性, P_j^s 与 P_j^N 完全正交,因此

$$I_j^T \cdot P_j^N(k) = 0, k = K+1, \dots, n \quad (3)$$

选取所有的噪声特征矢量作为水印信号集,即 $S_j = \{P_j(k), k=K+1, \dots, n\}$, 其中共有 $n-K$ 个水印信号,每个水印信号对应一个维数为 n 的匹配滤波器。显然 n 越大,匹配效果越好。 $n-K$ 可表示字符集的大小。 $n-K$ 越大,意味着有意义水印的内容越丰富。 I_j 作为一个信号矢量,其协方差矩阵 R_j 的秩仅为 1,因此信号特征值的个数 $K=1$ 。对每个矢量重复相同的操作,则可以得到 M 个类似的信号集。因此所需的特征集 $\{I_j\}$ 长度是 Mn 。可以看出, M 与 n 均不能无限增大,其乘积要受到图象实际特征集长度的制约。

式(3)意味着水印信号可以完全去除主数据干扰。无论水印信号的维数 n 为多大,根据特征子空间分解的特性,其中任意两个信号之间一定满足正交性。以随机产生的伪随机序列或二值序列作为信号集,不可能具有这一性质,而且信号集 S_j 是基于原始图象而产生的,具有一定程度的不可逆性,因此更加有利于解决版权问题^[4,5]。取 $n=96$, 由于 $K=1$, 信号集中的入口个数为 95,基本上可以映射 ASCII 码表中的所有字符,从而满足嵌入任意字符串的需要。利用信号集的入口对有意义信息进行编码,即可得到有意义水印序列 $\{w_i, i=1, \dots, Mn\}$ 。

2 基于小波域的水印嵌入算法

鉴于小波域水印良好的稳健性和感知质量^[3],特征集 $\{I_i\}$ 由小波系数构成。在没有原始图象的情况下,由测试图象得到的特征集 $\{X_i\}$ 可能与 $\{I_i\}$ 不完全相同。为确定特征集 $\{I_i\}$ 中的正确系数位置,水印嵌入分为两种情况:一是利用固定区域的系数作

为特征集,由于低频系数具有较好的抗压缩稳健性,选取低频子带 LH_3, HL_3, HH_3 中的所有系数作为特征集;二是选用 M_n 个最大的系数,也即将感知最重要的系数作为特征集^[2],把特征集 $\{I_i\}$ 中的系数位置作为附加信息在水印提取过程中使用,此时附加信息同样可以作为密钥,以保证水印的安全性. 水印嵌入公式如下:

$$I_i^w = I_i + \alpha w_i \quad (4)$$

其中, α 是水印信号强度因子, I_i^w 是修改过的特征系数,利用式(4)更易于保持水印信号之间的正交性. 由于水印信号集是由归一化的噪声特征矢量构成,每个水印信号 $P_j(k)$ 均值近似为 0,因此其方差为

$$\sigma_p^2 = E(P_j^2(k)) \approx \frac{1}{n} P_j(k)^T P_j(k) = \frac{1}{n} \quad (5)$$

可见,水印信号的能量远小于通常的高斯噪声水印^[2],因此可以通过相应地调整强度因子 α 来保证水印的稳健性.

3 基于匹配滤波的水印盲提取算法

给定待测试的图象,在情况 1 时, $\{X_i\}$ 可由固定区域及密钥 Key 获得. 在情况 2 时,要获得特征集 $\{X_i\}$ 必须同时拥有密钥 Key 和地址附加信息. 对特征集 $\{X_i\}$ 进行相应处理,得到矢量 $X_j, j=1, \dots, M$. 译码时把 X_j 与相应信号集 S_j 中所有水印信号进行匹配滤波,选取具有最大相似度的水印信号入口作为提取的字符. 相似度函数为^[3]

$$Sim(X_j, P_j(k)) = \frac{X_j^T \cdot P_j(k)}{\sqrt{X_j^T \cdot X_j}}, k=K+1, \dots, n \quad (6)$$

其中, $X_j = I_j + \alpha P_j(k_0) + N_j, k_0$ 是嵌入字符的入口, N_j 是迭加的噪声. 根据式(3)以及水印信号之间的正交特性,式(6)可以写成

$$Sim(X_j, P_j(k)) = \begin{cases} \frac{\alpha + N_j^T \cdot P_j(k_0)}{\sqrt{X_j^T \cdot X_j}}, & k = k_0 \\ \frac{N_j^T \cdot P_j(k)}{\sqrt{X_j^T \cdot X_j}}, & k \neq k_0 \end{cases} \quad (7)$$

给定 X_j , 分母 $\sqrt{X_j^T \cdot X_j}$ 对任意 k 均为常数. 当强度因子 α 足够大时,就可以准确提取字符入口 k_0 . 当噪声 N_j 为 0 时,可以得到理想值

$$Sim(X_j, P_j(k)) = \begin{cases} \frac{\alpha}{\sqrt{X_j^T \cdot X_j}}, & k = k_0 \\ 0, & k \neq k_0 \end{cases} \quad (8)$$

4 置信度门限

如何评估所提取的字符是否准确可靠,即置信度,这个工作往往被其他文献忽视. 在有意义水印的提取过程中,存在假设检验

$$\begin{aligned} H_0: X_j &= I_j + N_j, \text{ 不存在水印} \\ H_1: X_j &= I_j + \alpha P_j(k_0) + N_j, \text{ 存在水印} \end{aligned} \quad (9)$$

其中, N_j 表示各种图象处理操作或攻击产生的失真噪声. 在理想情况下,即不存在噪声,根据式(8),水印盲提取可使虚警概率和漏警概率均为 0. 下面讨论噪声不为 0 的情况. 根据有意义水印的盲提取算法,在假设 H_0 时,始终会存在某个 k 可得到最大的相似度输出,从而得到并不存在的有意义水印,即虚警概率等于 1. 显然在基于匹配滤波的有意义水印中,这是不可避免的现象,并使人对算法可靠性产生质疑.

在假设 H_1 时,如果存在某个 k 满足

$$Sim(X_j, P_j(k)) > Sim(X_j, P_j(k_0)) \quad (10)$$

这时会提取错误的字符入口,而漏掉真正的入口 k_0 . 当噪声干扰较大时,嵌入的水印信号与其他水印信号之间的正交性被破坏,就很可能出现式(10)的情况. 这种错误概率可以看作是漏警概率. 出现虚警概率和漏警概率都是很严重的问题,所以设置一置信度门限

是必要的. 由式(7)可以看出,分母 $\sqrt{X_j^T \cdot X_j}$ 可以看作是常数. 对任意 $k \neq k_0$, 其相似度主要取决于分子 $N_j^T \cdot P_j(k)$. 由于 $P_j(k)$ 是相互正交的噪声特征矢量,可以假设其所有分量均服从均值 0 的正态分布,且相互独立. 给定测试图象,此时噪声可以看作是常数序列,根据中心极限定理可知, $N_j^T \cdot P_j(k)$ 服从以下正态分布^[2],即 $N_j^T \cdot P_j(k) \sim N(0, N_j^T \cdot N_j \cdot \sigma_p^2)$. $N_j^T \cdot P_j(k)$ 的方差很小. 当 N_j 服从标准正态分布时,由式(5)可以得到 $N_j^T \cdot P_j(k)$ 的方差为 1. 由于分母 $\sqrt{X_j^T \cdot X_j}$ 远大于 $N_j^T \cdot P_j(k)$ 的方差,因此

$$Sim(X_j, P_j(k)) \sim N(0, \sigma^2), k \neq k_0 \quad (11)$$

其中方差 $\sigma^2 = N_j^T \cdot N_j \cdot \sigma_p^2 / \sqrt{X_j^T \cdot X_j} \ll 1$, 而

$$Sim(X_j, P_j(k_0)) \sim N(m, \sigma^2), k = k_0 \quad (12)$$

其中, $m = \alpha / \sqrt{X_j^T \cdot X_j}$. 当 α 足够大时,就可以准确地提取嵌入的字符. 根据式(11)、式(12),似乎可以类似地设置一 γ 门限,当相似度大于这一 γ 门限时,则判断水印存在,反之则不存在^[2]. 但是这种门限在此不适用,原因有两个:(1)由每一段矢量 X_j 得到的相似

度输出受分母 $\sqrt{X_j^T \cdot X}$ 影响很大。 $\sqrt{X_j^T \cdot X}$ 受向量 I_j 和噪声 N_j 的影响而随 j 变化, 因此不存在一个统一门限; (2) 当噪声干扰较大时, 水印被破坏, 由幅值门限得到的入口 k 不具有可靠性。

为解决这一问题, 设置置信度 c_j

$$c_j = \begin{cases} 1, \max_k (Sim(X_j, P_j(k))) / \max_k (Sim(X_j, P_j(k))) \leq T \\ c_j, \max_k (Sim(X_j, P_j(k))) / \max_k (Sim(X_j, P_j(k))) > T \end{cases} \quad (13)$$

其中, T 是置信度门限, $\max(\cdot)$ 是取最大值函数, $\max'(\cdot)$ 是取次最大值函数。当 $c_j=1$ 时, 说明所提取的字符是可靠的; 反之则不可靠。当噪声干扰较小时, $Sim(X_j, P_j(k_0))$ 远大于其他入口的相似度输出; 当噪声干扰太大或根本就没有水印时, 最大相似度输出随机而定, 且与其他相似度输出相差不大。式

(13) 有效地反映了这一关系。

可见, 门限 T 越小, 虚警概率和漏警概率越低, 但也可能导致提取的字符入口正确, 而其置信度仍为 0。可是, 由于有意义水印算法的可信赖程度是推动其应用的重要因素, 因此水印算法设计的重点应放在可靠性上。从这一意义上, 保守的估计与较高的虚警概率和漏警概率相比, 显得更为重要。为评估整体可靠性, 整个水印字符串的置信度 $Cred$ 可以表示为

$$Cred = \frac{1}{M} \sum_{j=1}^M c_j \quad (14)$$

值得注意的是, 当整体水印字符串的置信度 $Cred$ 较大时, 有意义的水印字符串可以天然地作为纠错码, 从而对少数置信度为 0 的字符进行纠正。例如字符串“watormark”, 显然可以纠正为“watermark”。



(a) Lena 图象 (b) 固定区域加水印图象 (PSNR=51.3dB) (c) 最大系数加水印图象 (PSNR=42.9dB)

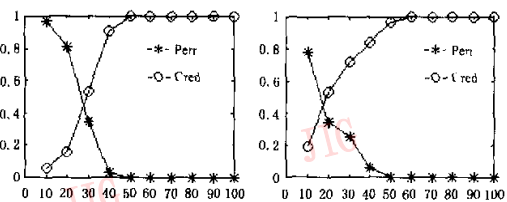
图 1

5 实验结果

以 $256 \times 256 \times 8$ 的标准 Lena 图象作为原始图象, 水印按照两种情况嵌入。对固定区域 LH_3, HL_3, HH_3 取强度因子 $\alpha=29.4$, 对最大系数特征集取强度因子 $\alpha=78.4$ 。所需的特征集必须利用密钥 Key。实验中, 特征集的长度均为 3072, 取 $n=96$, 总共可以嵌入 $M=32$ 个字符。图 1 分别给出了原始图象和加水印图象。两幅加水印图象均无可感知的失真。固定区域加水印图象 (图 1(b)) 峰值信噪比 (PSNR=51.3dB), 要大于最大系数加水印图象 (图 1(c)) (PSNR=42.9dB), 这是因为其强度因子较小, 嵌入的水印能量也较小的缘故。虽然嵌入水印的能量不相同, 两种加水印方法对不同失真的稳健性各有胜出。

取置信度门限 $T=3/4$ 。图 2(a)、(b) 分别给出了图 1(b)、(c) 在 JPEG 压缩下的稳健性测试结果。

从测试结果可见, 二者相差不大, 当 JPEG 压缩质量因子 Q 大于 50 时, 均可达到误码率为 0, 但是固定区域水印, 在 $Q=50$ 时, 置信度为 1, 即提取的字符完全可靠; 而最大系数所加水印, 在 $Q=50$ 时, 置信度为 0.9688, 即有 1 个字符的可靠性值得怀疑。这说明固定区域水印的稳健性略强于最大系数所加水印, 其原因是 JPEG 压缩对低频区域 LH_3, HL_3, HH_3 保存较好。



(a) 固定区域水印测试结果 (b) 最大系数加水印图象测试结果

图 2 JPEG 压缩下误码率 $Perr$ 和置信度曲线 $Cred$

