



# 图像信息隐藏算法三

---

钮心忻、杨榆、雷敏

北京邮电大学信息安全中心

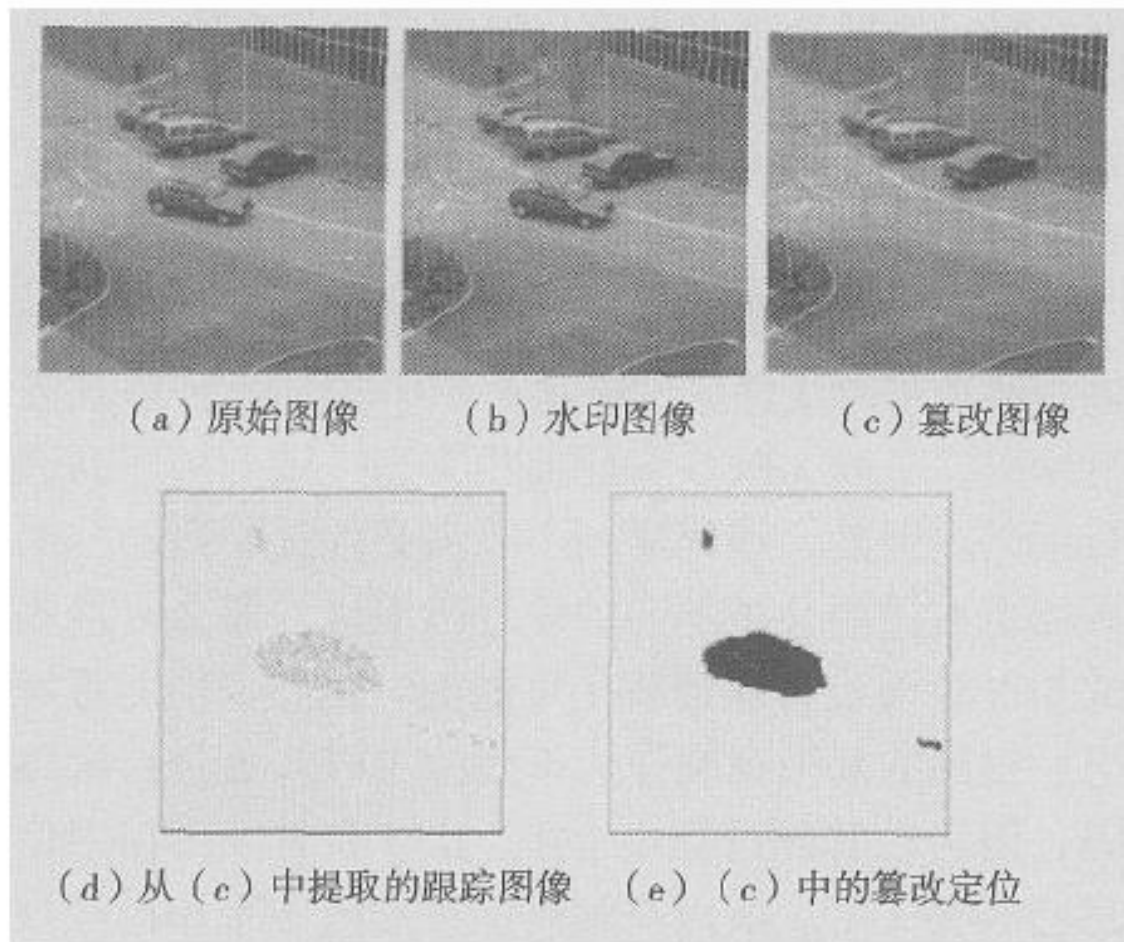
[yangyu@bupt.edu.cn](mailto:yangyu@bupt.edu.cn)

# 图像水印算法介绍

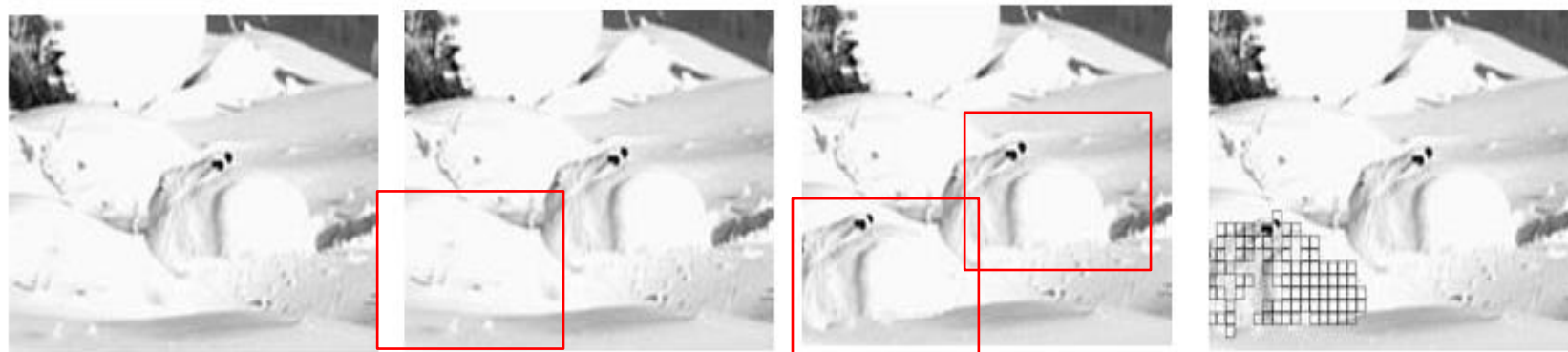
---

- 普通图像水印
  - 隐蔽性
- 图像鲁棒性水印
  - 能够抵抗各种信号处理、攻击
- 图像脆弱性水印
  - 完整性验证、篡改定位

# 图像（半）脆弱水印的应用



# 图像（半）脆弱水印的应用



(a) 原始图像

(b) 水印图像

(c) 被篡改的图像

(d) 篡改部位标注



(e) 水印



(f) 从图(b)提出的水印



(g) 从图(c)提出的水印



(h) 图(g)与(f)的差

# 图像（半）脆弱水印的应用



（a）块B1被压缩后隐藏到块B2的LSB中

## 图像（半）脆弱水印的应用

---



(b) 车牌部分被调换



# 图像（半）脆弱水印的应用

---



（c）提取并恢复后的图像（车牌被改正过来）

# 内容认证

---

## ○ 内容认证研究以下几个问题

- 作品是否被改变?
- 作品是否被显著改变?
- 作品哪一部份被改变?
- 被改变的作品能复原吗?



# 内容认证

---

- 基于密码学的完整性认证方案
  - 使用哈希函数生成内容的摘要。
  - 使用发送方私钥对摘要签名。
  - 发送内容及其签名到接收方。
  - 接收方验证签名，判定内容完整性。
- 每一个问题都有水印以外的解决方案，水印解决方案的特点在于：
  - 不需要额外的数据存储认证信息。
  - 传输过程中，水印经受与载体相同的处理。

# 内容认证

---

## ○ 完全内容认证

- 完全内容认证系统用于验证作品是否一点没变

## ○ 主要用途

- 医学图像，法庭证物

## ○ 主要方法

- 脆弱水印：算法被设计为，即使作品仅被微小改变，水印也会消失。
- 嵌入签名：用密码技术产生的签名作为水印嵌入到载体中。

# 内容认证——完全内容认证

---

## ○ 基于脆弱水印的认证方法

- 使用LSB将水印嵌入作品。
- 传输过程中，若有噪声、滤波、压缩编码等等环节，水印将消失。
- 这样的作品不能通过接受方的认证。
- 特点：
  - 水印与载体无关。

# 内容认证——完全内容认证

---

- 基于脆弱水印的认证方法的安全问题
  - 可以搜集多幅水印图像，拼凑出篡改图像，同时能通过认证。
  - 可以修改非水印区域同时通过认证。

# 内容认证——完全内容认证

## ○ 基于脆弱水印的认证方法的安全问题



图 (a) 测试图像



图 (b) 水印



图 (c) 水印图像



图 (d) 攻击图像



图 (e) 被篡改的水印图像



图 (f) 提取出的水印

# 内容认证——完全内容认证

---

## ○ 嵌入签名的方法

- P. W. Wong 水印系统
- 基于公钥的图像认证和完整性数字水印系统。
- 利用hash函数的单向性和“雪崩效应”定位篡改。
- 借助公钥系统的便利性，公钥的用户完成对图像的完整性检测和身份认证。

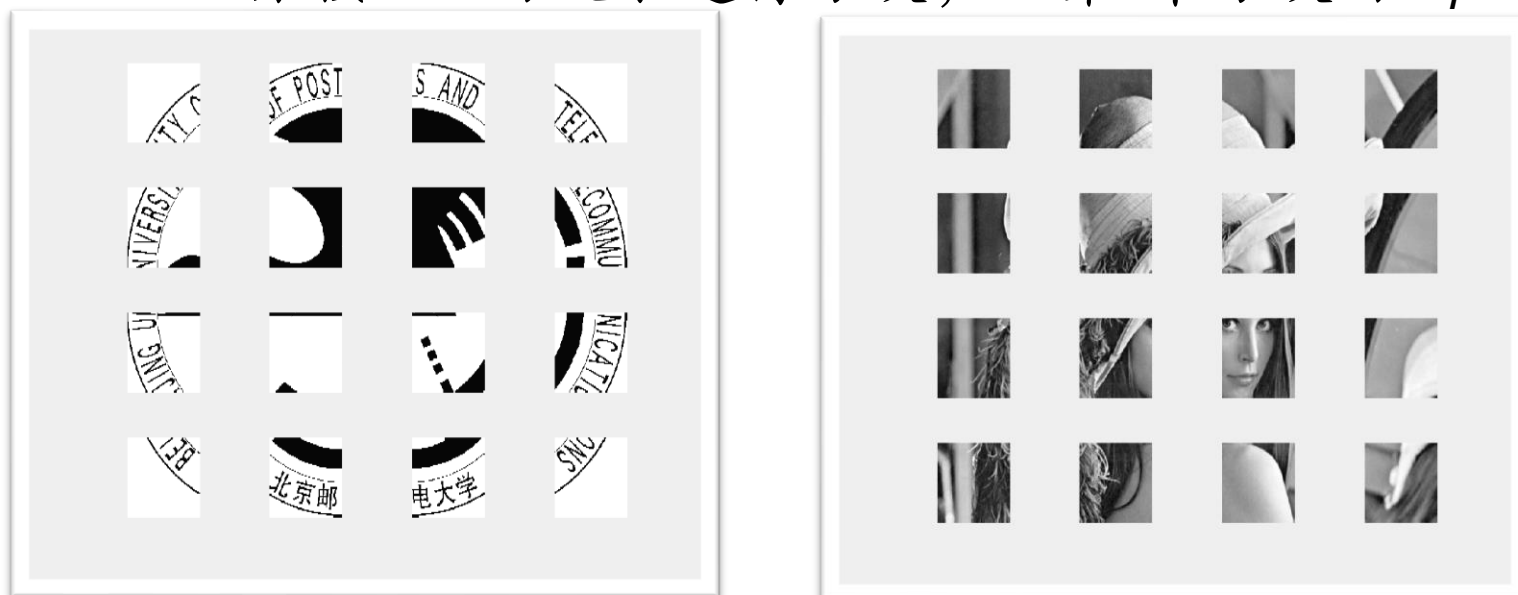


# 内容认证——

## P. W. Wong 水印算法的流程

### ○ 水印嵌入过程

- 将原始图像( $M \times N$ 大小)按 $I \times J$ 的大小进行分块, 记作品第 $r$ 个分块为 $X_r$ 。
- 将水印图像扩充为与原始图像同样大小的尺寸, 亦按 $I \times J$ 的大小进行分块, 记第 $r$ 个分块为 $W_r$ 。



# 内容认证——

## P. W. Wong 水印算法的流程

---

### 水印嵌入过程

- 将作品分块 $X_r$ 的LSB 置为零，得到主信息分块 $X'_r = \text{And}(X_r, (0xFE)_{I \times J})$ ;
- 以主信息分块 $X'_r$ 和M,N 为参数，计算摘要，得到内容的“指纹”分块 $P_r = H(X'_r|M|N)$ ;
- 按位异或指纹分块 $P_r$ 与水印分块 $W_r$ ，结果用公钥系统加密（用作者的私钥），得到认证信息 $A_r = C(P_r \oplus W_r, K_{pri_A})$ ;
- 将认证信息置入主信息分块 $X'_r$ 的最低有效位，得到认证分块 $Y_r = OR(A_r, X'_r)$ 。
- 重组所有认证分块，生成嵌入水印后的图像，认证作品 $Y = (Y_r)$

# 内容认证——问题

---

- 接收端如何完成图像认证？
- 与基于LSB的脆弱水印方案的核心区别是什么？替换非水印区域比特是否能够通过认证？
- 不使用水印图像能否完成认证？水印图像的用途是？
- 为什么要选用非对称密钥系统？

# 内容认证——

## P. W. Wong 水印算法的流程

### 水印提取和验证过程

将水印图像( $M \times N$ )按 $I \times J$ 大小分块, 记第 $r$ 个分块为 $Y_r$ ;

将水印扩充, 使其与水印图像的大小相同, 亦按 $I \times J$ 的大小进行分块, 记第 $r$ 个分块为 $W_r$ ;

- 将水印作品分块 $Y_r$ 的LSB 置零, 得到的分块记为主信息分块 $Y'_r = \text{And}(Y_r, (0xFE)_{I \times J})$ ; 若无篡改, 水印作品主信息分块 $Y'_r$ 和原作品主信息分块 $X'_r$ 相同;
- 以主信息分块 $Y'_r$ 和 $M, N$  为参数, 计算得到指纹分块 $P_r = H(Y'_r | M | N)$ ;
- 按位异或指纹分块 $P_r$ 与水印分块 $W_r$ , 得到验证块1,  $PW1_r = P_r \oplus W_r$ ;
- 取出分块最低位平面, 并用作者公钥解密, 得到验证块2,  
 $PW2_r = D(\text{And}(Y_r, (0x01)_{I \times J}), K_{pub_A})$ ;
- 若 $PW1_r$ 与 $PW2_r$ 相等, 则说明该块没有被篡改, 通过认证;
- 依次处理所有分块, 就能判定图像是否被篡改。

# 内容认证——

## P. W. Wong 水印算法的安全性分析

---

○ 与基于LSB的脆弱水印方案的核心区别是什么？

- 认证信息关联了作品内容和水印

$$PW1_r = P_r \oplus W_r = H(\text{And}(X_r, (0xFE)_{I \times J}) | M | N) \oplus W_r$$

- 认证信息用公钥密码体制保护

$$A_r = C(P_r \oplus W_r, K_{pri_A})$$

# 内容认证——

## P. W. Wong 水印算法的安全性分析

### ○ 替换非水印区域比特是否能通过认证？

- 不能。替换非水印区域将导致基于内容计算的验证信息变化，使之有别于基于原作品生成的验证信息。

$$PW1_r = P_r \oplus W_r = H\left(\text{And}(Y_r^p, (0xFE)_{I \times J}) | M | N\right) \oplus W_r$$

$$PW2_r = D\left(\text{And}(Y_r^p, (0x01)_{I \times J}), K_{pub_A}\right) = H\left(\text{And}(X_r, (0xFE)_{I \times J}) | M | N\right) \oplus W_r$$

### ○ 不使用水印图像能否完成认证？ 水印图像的运用是？

- 无水印图像破坏认证的完整性。认证算法的操作粒度是尺寸为  $I \times J$ 。无水印图像辅助，可以扰乱同一水印作品内图像小块的顺序，或者交换不同作品间小块，同时通过认证。

### ○ 为什么要选用非对称密钥系统？

- 选用非对称密码体制保护认证信息，认证系统机制可以公开同时能防范攻击者伪造认证信息。



# 内容认证

---

## ○ 选择认证系统

- 预定义引入合法失真的处理集，和引入非法失真的处理集，当作品经受前者而没有后者处理时，能够通过系统认证。

## ○ 有三类基本方法：

- 半脆弱水印：遭遇合法处理时，水印能够生存，遭遇非法处理时，水印消失。
- 半脆弱签名：以不受合法处理影响，但遭受非法处理时会发生改变的载体的特征为水印嵌入，水印算法可选稳健算法或半脆弱算法。

# 内容认证

---

- 有三类基本方法：
  - Telltale水印：用于研究载体经受哪些操作。通过研究水印的变化，推断载体经受哪些操作，最终判断载体是否能通过认证。

# 内容认证——半脆弱水印算法

---

- 半脆弱水印主要有以下思路：
  - 与JPEG相结合的半脆弱数字水印算法：
    - 这类水印认证算法一般是根据JPEG编、解码器的特点而设计的, 故水印算法通常对JPEG压缩具有较好的鲁棒性, 而对其他的图像操作反应敏感
  - 从鲁棒水印算法演变而来的半脆弱水印算法：
    - 该类算法主要是借鉴鲁棒图像水印算法的一些经典方法(如扩频水印)来设计相应的认证算法

# 内容认证——半脆弱水印算法

---

- 基于视觉掩蔽模型的半脆弱水印算法
  - 将人眼视觉掩蔽模型应用于数字水印系统，会使嵌入水印后的图像具有更好的主观视觉质量
- 基于小波域的半脆弱水印算法
  - 由于小波变换是一种空间—频率分析方法，能同时反映图像的空间位置和频率
  - 小波变化的局部化作用能够检测到图像被篡改的区域
  - 小波变化的频率域则反映了被篡改的尺度（频带）








# 内容认证——半脆弱水印








---

- 二值图像作为水印
- 音频分段做DWT和DCT变化。
- 量化系数嵌入水印，每次嵌入一行。
- 系数做IDCT和IDWT变化。
- 重组所有分段。
- 音频分段做DWT和DCT变化。
- 提取水印 $W'$ 。
- $W'$ 与原始水印 $W$ 做抑或。
- 判决载体是否被篡改。

# 内容认证——半脆弱水印

## ○ 认证效果（图片、数据源自参考文献2）

		篡改	篡改+ 重新采样	篡改+ 重新量化	篡改+ 加白噪声	篡改+ MP3-64kb	篡改+ MP3-56kb	篡改+ 低通滤波
本文算法	水印图像							
	NC	0.8469	0.7803	0.8466	0.8444	0.8481	0.8492	0.8471
	PSNR	30.705	27.6016	30.5945	30.5807	30.7803	30.7932	30.1195

	含水印音频		常规攻击		恶意篡改	联合攻击	
	未攻击	重新采样	MP3-56 kb	篡改	篡改+ 重新采样	篡改+ MP3-56 kb	
篡改矩阵 (已去噪)							
$R_{num}$	0	0	0	12	12	12	
$R_{total}$	0	0	0	0.1875	0.1875	0.1875	



# 内容认证——半脆弱签名

## ○ 基于JPEG的半脆弱签名算法

### ● 原理：

- 定理1：系数的大小相对关系经过均匀量化能够得到保存
- 定理2：均匀量化时，如果系数首先被一个较大的量化步长量化，则能够使用较小的量化步长将系数无损的恢复出来。
$$x \circ q = q \left\lfloor \frac{x}{q} + 0.5 \right\rfloor$$
- 若定义量化算子如下
- 则定理2可以表示为： $(q_2 \leq q_1)$

$$((a \circ q_1) \circ q_2) \circ q_1 = a \circ q_1$$

# 内容认证——半脆弱签名

---

## ○ 基于JPEG的半脆弱签名算法

- 根据定理1，可以产生签名。
  - 把图像分块两两配对，比较每对特定位置DCT系数大小，生成签名。
- 根据定理2，可以嵌入签名。
  - 用较大的量化步长，把签名嵌入量化系数的最低比特。
  - 当图像经受JPEG压缩时，根据定理2可知，只要量化步长小于嵌入签名时使用的步长，签名信息都能正确地恢复。

# 内容认证——半脆弱签名

○ 认证效果（图片、数据源自参考文献3）



Figure 2: (a) The original image, (b) the watermarked image after embedding authentication bits ( PSNR = 40.7 dB), (c) the watermarked image after embedding authentication bits and weak recovery bits ( PSNR = 37.0 dB)

# 内容认证——半脆弱签名

## ○ 认证效果（图片、数据源自参考文献3）

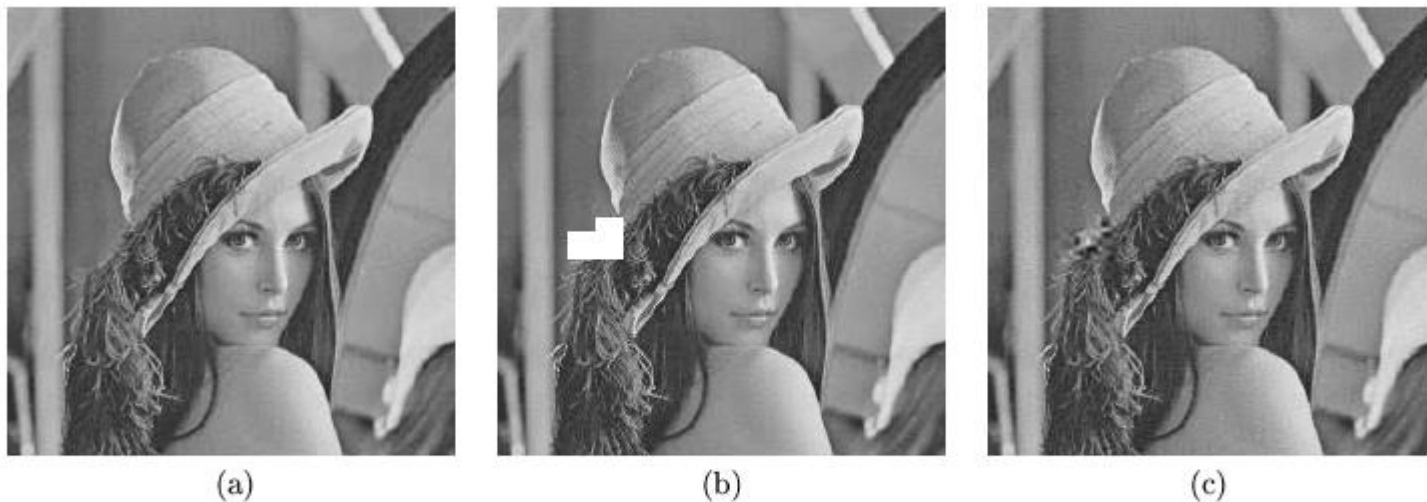


Figure 3: (a) Manipulation on the watermarked image in Figure 2(b), (b) the authentication result of (a), (c) the authentication result from the manipulated image of Figure 2(c)

## 参考文献

---

1. Ingemar J Cox, Mathew I Miller, Jeffrey A Bloom, Jesscia Fridrich, Ton Kaller. Digital Watermarking and Steganography.
2. 王向阳, 祁薇。用于版权保护与内容认证的半脆弱音频水印算法。
3. Ching-Yung Lin, Shih-Fu Chang, Semi-Fragile Watermarking for Authenticating JPEG Visual Content.