《数字内容安全》

课程讲义

第4讲 信息隐藏与数字水印

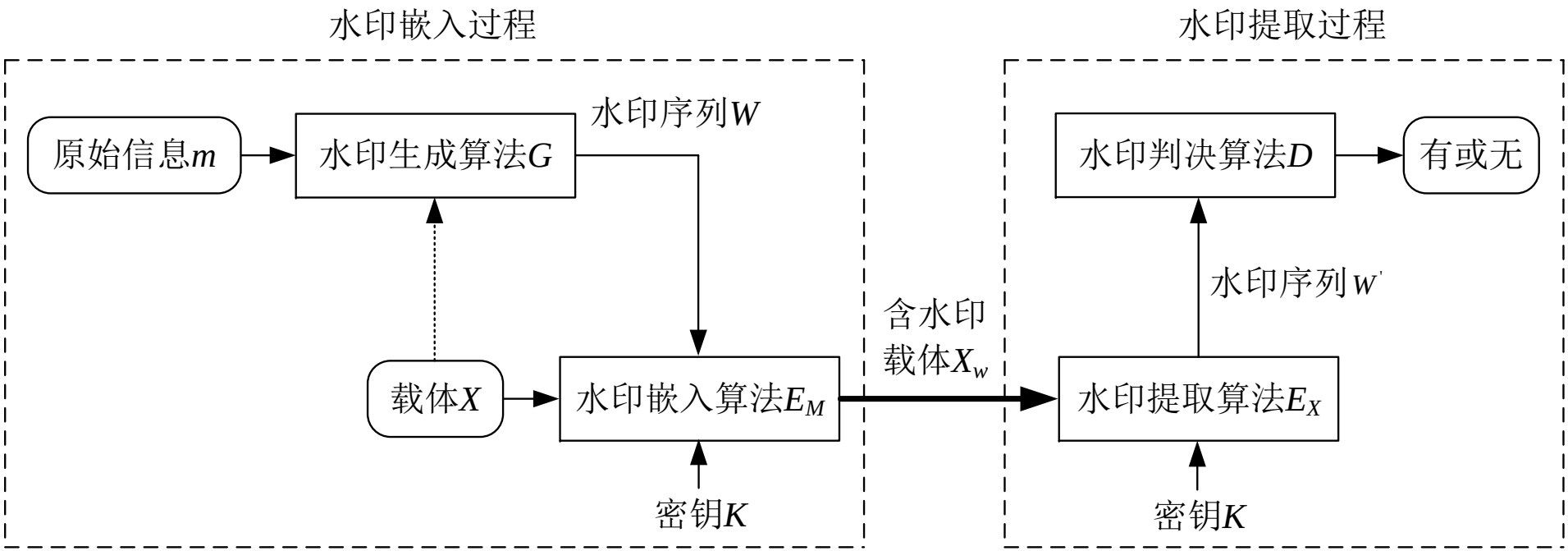
主要内容

1. 信息隐藏基本理论
2. 空域/变换域信息隐藏技术
3. 数字水印
4. 应用与发展

3 数字水印技术

- 框架
- 分类
- 评价指标
- 攻击方法
- 版权保护水印
- 内容认证水印
- 可逆水印

数字水印技术框架


水印嵌入过程
水印提取过程
原始信息m
水印生成算法G
水印序列W
载体X
水印嵌入算法E_M
密钥K
含水印
载体X_w
水印提取算法E_X
水印判决算法D
有或无
水印序列W'
密钥K

3 数字水印技术

- 框架
- 分类
- 评价指标
- 攻击方法
- 版权保护水印
- 内容认证水印
- 可逆水印

➢ 数字水印技术分类

（1）宿主信息：文本水印、图像水印、视频水印、矢量图水印等。

（2）水印嵌入位置：空域数字水印和变换域数字水印。

（3）数字水印的性质：鲁棒水印和脆弱水印。鲁棒水印主要用于数字内容信息的版权保护和所有权认定，故应能经受各种可能的攻击；脆弱水印可以进一步分为完全脆弱水印和半脆弱水印。

➢ 数字水印技术分类

（4）水印检测是否需要原始载体信息和水印信息：盲检测水印和非盲检测水印。

（5）检测方法的角度：水印可以分为私有水印和公开水印。

（6）水印载体是否能够无损恢复：可逆水印和不可逆水印。

3 数字水印技术

- 框架
- 分类
- 评价指标
- 攻击方法
- 版权保护水印
- 内容认证水印
- 可逆水印

➢ 数字水印技术评价指标

（1）安全性：水印的信息应是安全的，难以篡改或伪造。

（2）不可见性：数字水印应是不可知觉的，而且应不影响被保护数据的正常使用。衡量隐蔽性的客观标准有均方误差 MSE（mean-square error）和信噪比SNR（Signal-to-noise radio）和峰值信噪比PSNR。

$$MSE = \frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}\left|H'(x,y) - H(x,y)\right|^2$$

➢ 数字水印技术评价指标

（3）鲁棒性：指在经历多种有意或无意的信号处理操作后，数字水印仍能保持部分完整性并能被准确鉴别。

（4）水印容量：嵌入的水印信息必须足以表示数字内容的创建者或所有者的标志信息，或购买者的序列号，这样有利于保护数字产权合法拥有者的利益。

➢ 数字水印技术评价指标

Robustness criteria

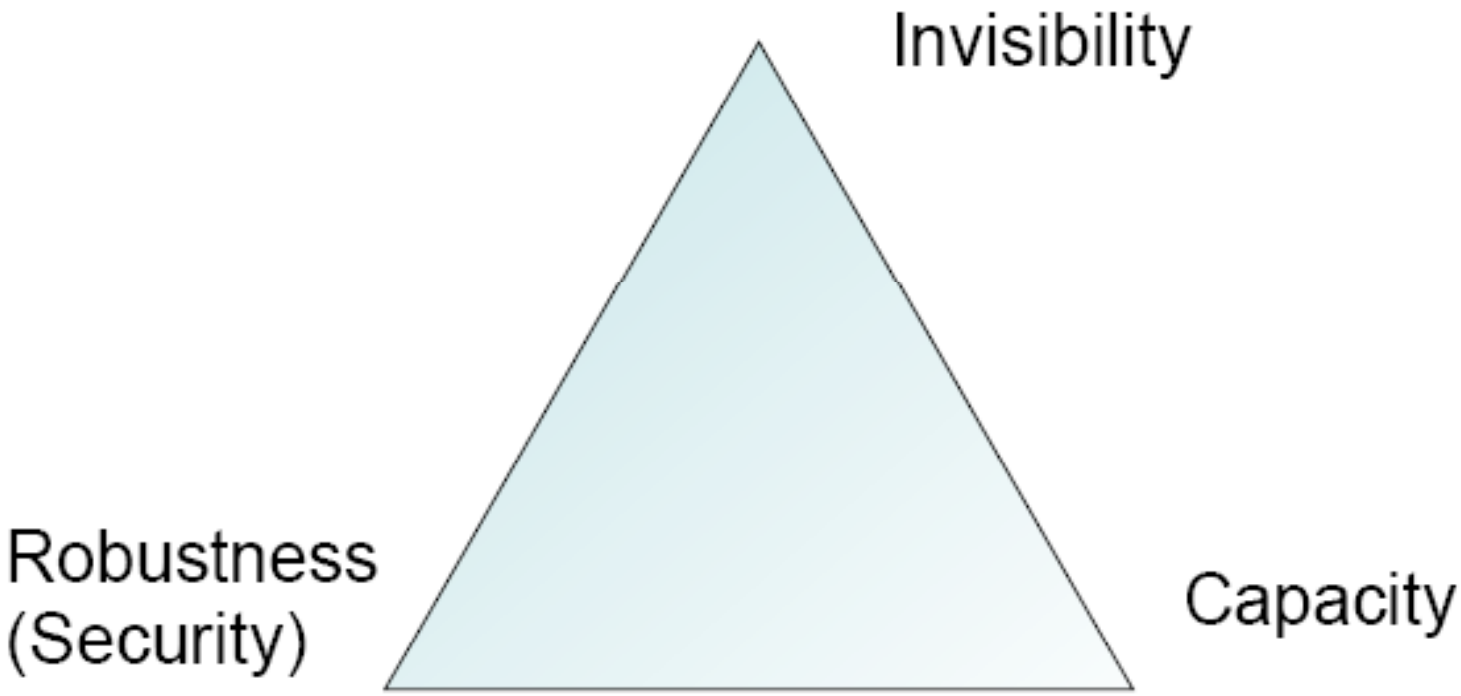
- Signal processing
  - enhancement, sharpening, blurring, linear/non-linear filtering (median, de-speckle)
- Compression
  - Robustness against JPEG compression is mandatory
- Geometric manipulations
  - resizing, cropping, translation, rotation, flip
- A/D – D/A conversion

➢ 数字水印技术评价指标

Robustness criteria

- For a quantitative measure of robustness, a given attack, or a combined attack, is carried out at different strength levels, the higher the level the watermark survives the higher the robustness
- Use of watermark-breaking tools
  - **Stirmark**: http://www.cl.cam.ac.uk/~mgk25/stirmark.html
  - **Certimark** - IST European Project: http://www.certimark.org

➢ 数字水印技术评价指标



3 数字水印技术

- 框架
- 分类
- 评价指标
- 攻击方法
- 版权保护水印
- 内容认证水印
- 可逆水印

➢ 数字水印的攻击方法

水印攻击与密码攻击一样，包括主动攻击和被动攻击。主动攻击的目的并不是破解数字水印，而是篡改或破坏水印，使合法用户也不能读取水印信息。而被动攻击则试图破解数字水印算法。

（1）解释攻击及对策

（2）信号处理攻击及其对策

（3）分析攻击及对策

（4）表达攻击及对策

3 数字水印技术

- 框架
- 分类
- 评价指标
- 攻击方法
- 版权保护水印
- 内容认证水印
- 可逆水印

版权保护数字水印技术

至今还没有一个数字水印的最终技术标准，但DHSG已经明确了用于版权保护的数字水印必须满足的一些基本条件：

（1）隐藏于数字作品中且不可感知；

（2）可以被专用的数字电路识别；

（3）不必获取完整数据，仅从数据流中即可检测到数字水印。

版权保护数字水印技术

（4）可以标记“未曾复制”、“只可复制一次”和“不能再复制”等复制信息；

（5）漏检概率低；

（6）水印内容（字段）的设计必须合理；

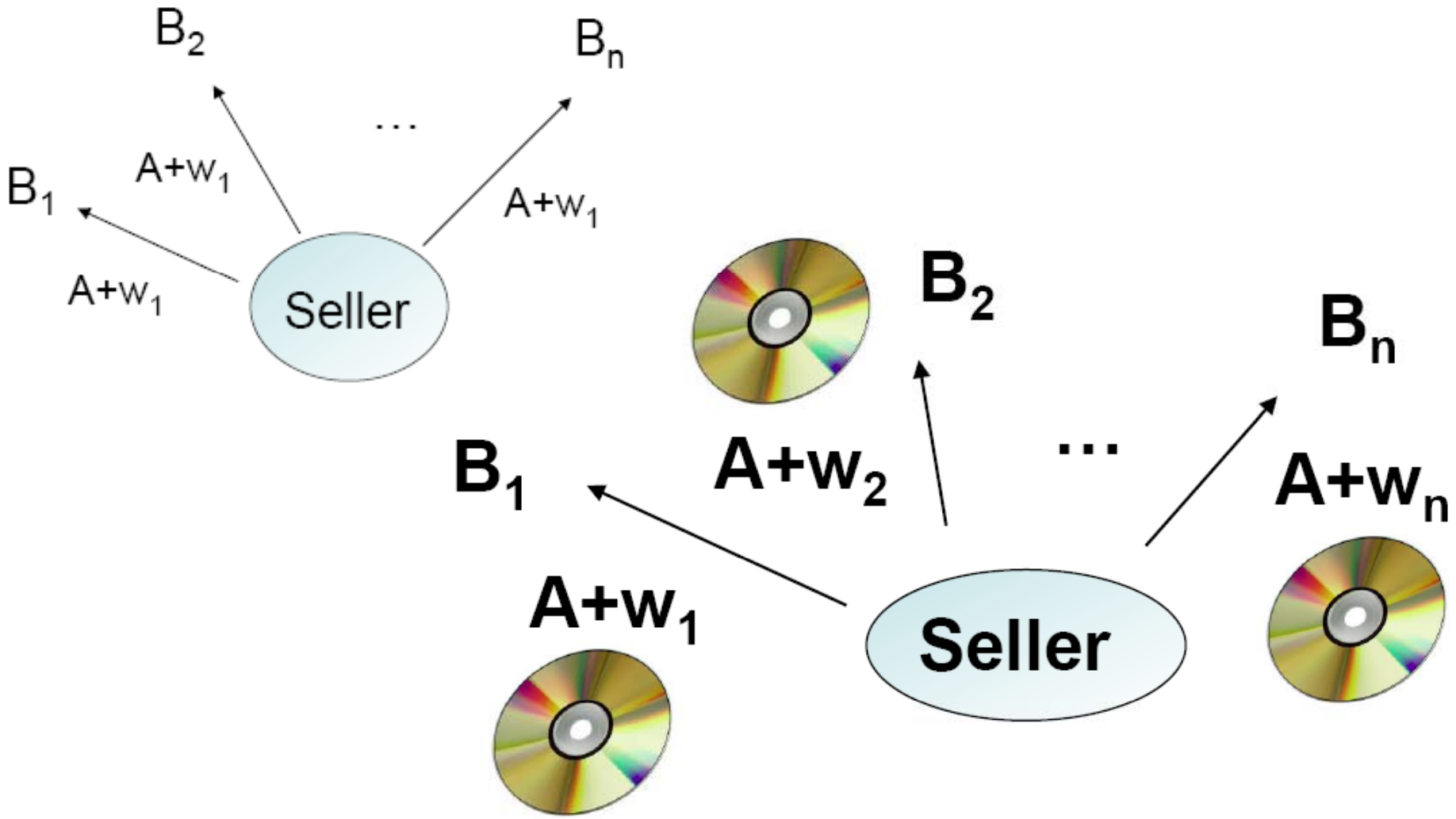
（7）必须使用成熟的技术嵌入或检测水印。

版权保护数字水印技术

Watermark content

- Depending on the application the information conveyed by the watermark may vary
  - Owner identification
  - Purchaser identification (fingerprinting)
  - Allowed uses
  - Transaction details
  - Authentication
  - Data labelling and tagging
  - Internet of things
  - ...

版权保护数字水印技术（Ownership Verification）



➢ 压缩域算法

基于JPEG、MPEG标准的压缩域信息隐藏系统不仅节省了大量的完全解码和重新编码过程，而且在数字电视广播及VOD（Video On Demand）中有很大的使用价值。相应地，水印检测与提取也直接在压缩域数据中进行。

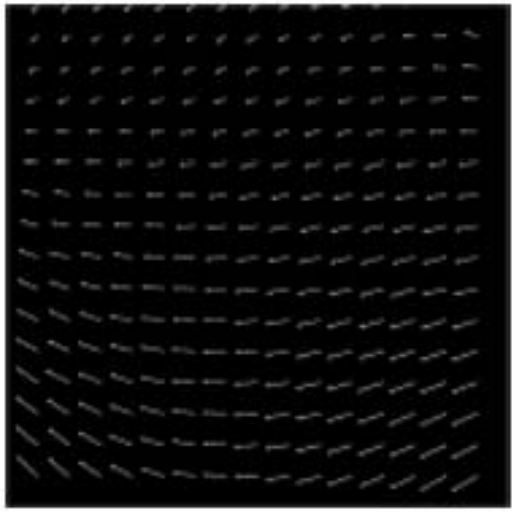
➢ RST域算法

RST域信息隐藏法的基本思想是利用Fourier-Mellin等变换，使得经过旋转、缩放、平移后得到的图像和原图像在RST域保持一致。此过程需要先后经过离散傅立叶变换、Fourier-Mellin变换、DFT。

该方法的优点是有很强的抗几何变换能力，缺点是抵抗有损压缩、低通滤波等信号处理方法的稳健性不够。

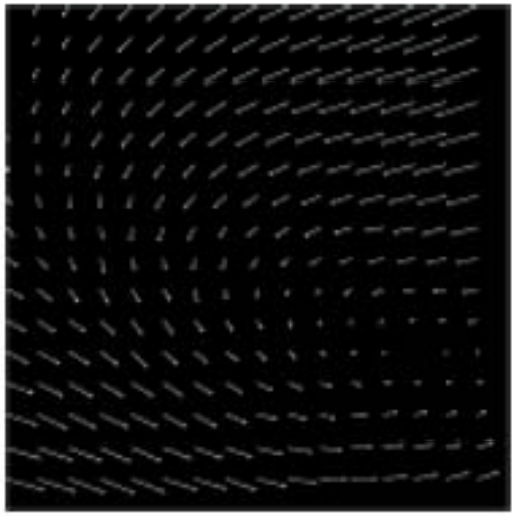
RBA: an example

Original



Stirmark 1

Example 1



Stirmark 2

Example 2

Multiresolution RBA

DF subsampled by 2

DF subsampled by 8

Multiresolution RBA

DF subsampled by 32

DF subsampled by 64

版权保护数字水印技术: Example 1

Example*

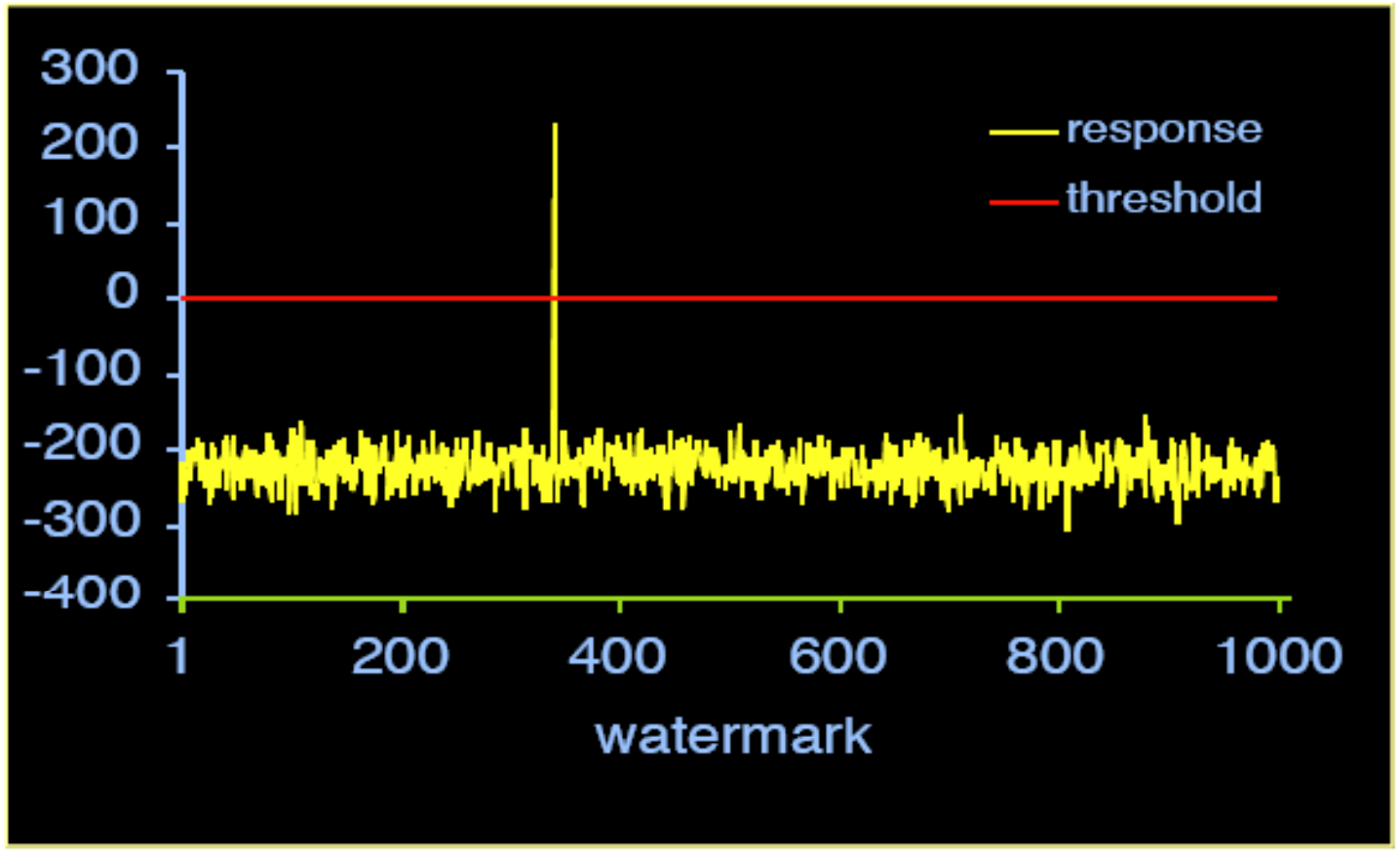
Original Image

Watermarked image
(PSNR = 50dB)

* M. Barni, F. Bartolini, A. De Rosa, and A. Piva, "A new decoder for the optimum recovery of non-additive watermarks, *IEEE Trans. Image Processing*, 10 (2001), pp. 755–766.

版权保护数字水印技术

Detector answer



版权保护数字水印技术

Example: robustness

JPEG compression with quality factor = 3%

版权保护数字水印技术

Example: robustness

Addition of white gaussian noise with variance = 2000

版权保护数字水印技术

Example: robustness

Print, copying and scanning

版权保护数字水印技术：Example 2

(a)
Original Lena

(b)
VQ compressed Lena, 31.53 dB

Fig. 2. The test image and VQ compressed one.

Cited from ‘VQ-Based Watermarking Techniques’

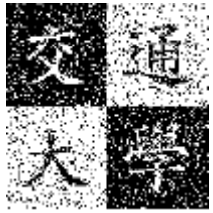
版权保护数字水印技术：Example 2

$NC_1 = 1.0$

(a) No attack

$NC_1 = 1.0$

(b) VQ, Codebook 1

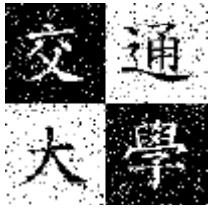


$NC_1 = 0.9707$

(c) VQ, Codebook 2

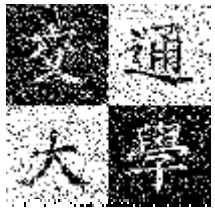
$NC_1 = 0.8869$

(d) VQ, Codebook 3



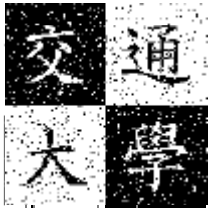
$NC_1 = 0.9888$

(e) JPEG, QF = 60%



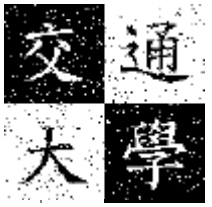
$NC_1 = 0.8604$

(f) Image cropping



$NC_1 = 0.9745$

(g) Low-pass filter



$NC_1 = 0.9848$

(h) Median filter

Fig. 3. The extracted watermarks with size 128×128 and the NC values of the proposed algorithm under various attacking methods.

Cited from ‘VQ-Based Watermarking Techniques’

3 数字水印技术

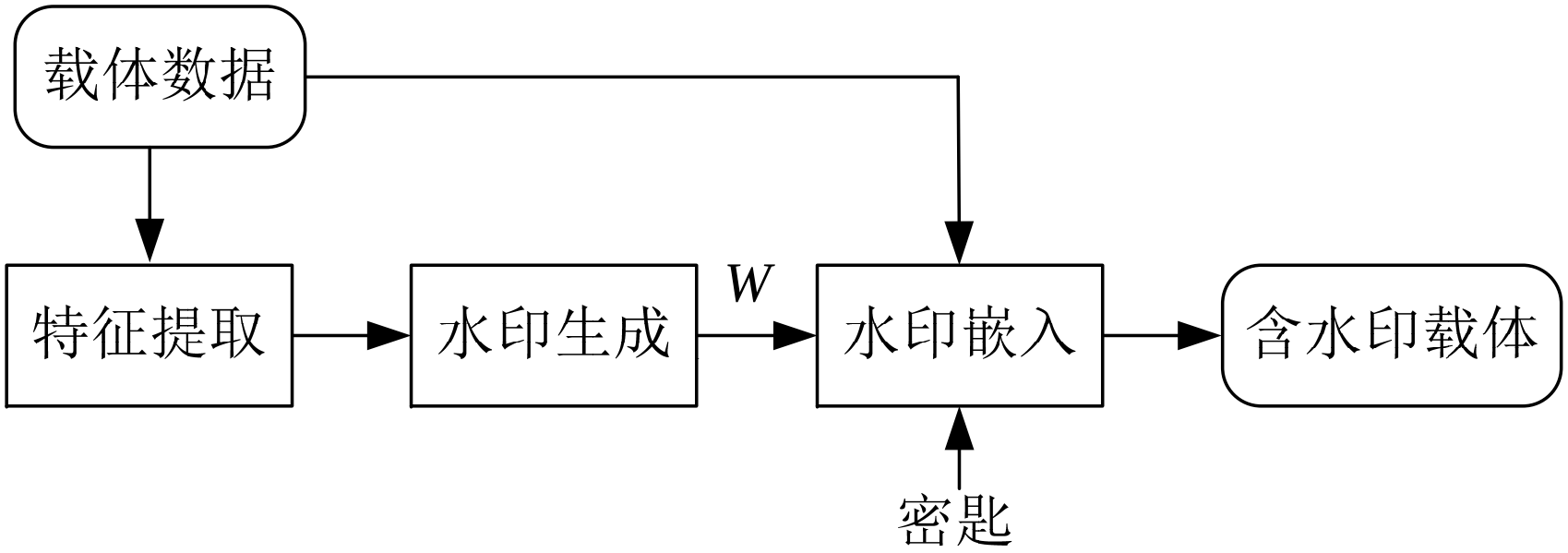
- 框架
- 分类
- 评价指标
- 攻击方法
- 版权保护水印
- 内容认证水印
- 可逆水印

➢ 内容认证水印技术的基本原理

基于数字水印的内容认证技术的基本原理是，预先在原始载体中通过水印嵌入器隐藏某种认证信号；在需要验证时，通过水印检测器识别这些认证信号的变动，以鉴别待测载体数据的完整性和真实性。通常，这里的水印信号可以是人为叠加的辅助信号或模板信号，也可以是依据载体内容或内容特征生成的校验信息。

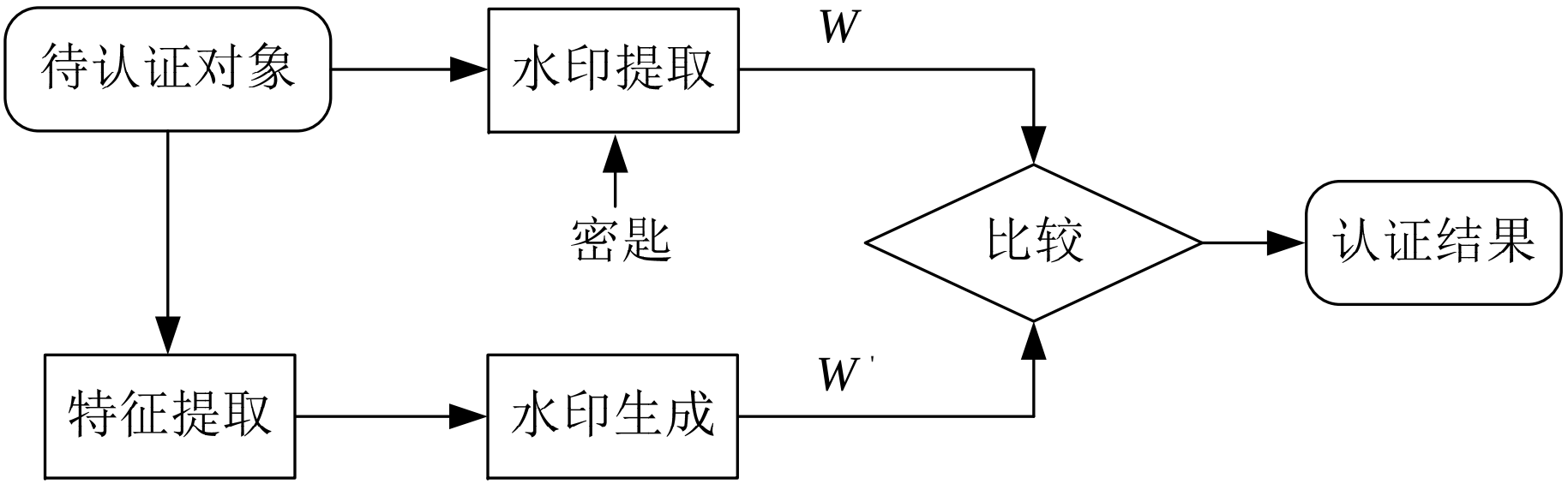
一个完整的数字水印认证系统通常分为水印嵌入和水印检测两个部分。

认证水印的嵌入过程



内容认证数字水印要检测出篡改位置并进行定位，其嵌入过程是：首先对原始载体进行特征提取并以此来构造水印信息，再将水印信息嵌入到原始载体中就得到嵌入水印后的受保护数字内容。

认证水印的检测过程



➢ 图像内容认证过程如图所示。认证时，根据密钥提取出受保护图像中的水印信息。然后将提取出来的水印信息W与重新生成的水印信息W‘相比较，若二者一致，则图像未被更改；若二者不一致，则认为图像已被改动，并给出关于改动的详细信息。

➢ 认证水印技术的要求

理想的图像认证系统应满足以下四项基本准则：

(1) **敏感性**：系统应敏感于改变了图像语义的恶意操作，如复制粘贴、拼接和剪切等；

(2) **容忍性**：系统应能忍受普通的信息损失和非恶意操作，如有损压缩、加噪和滤波等；

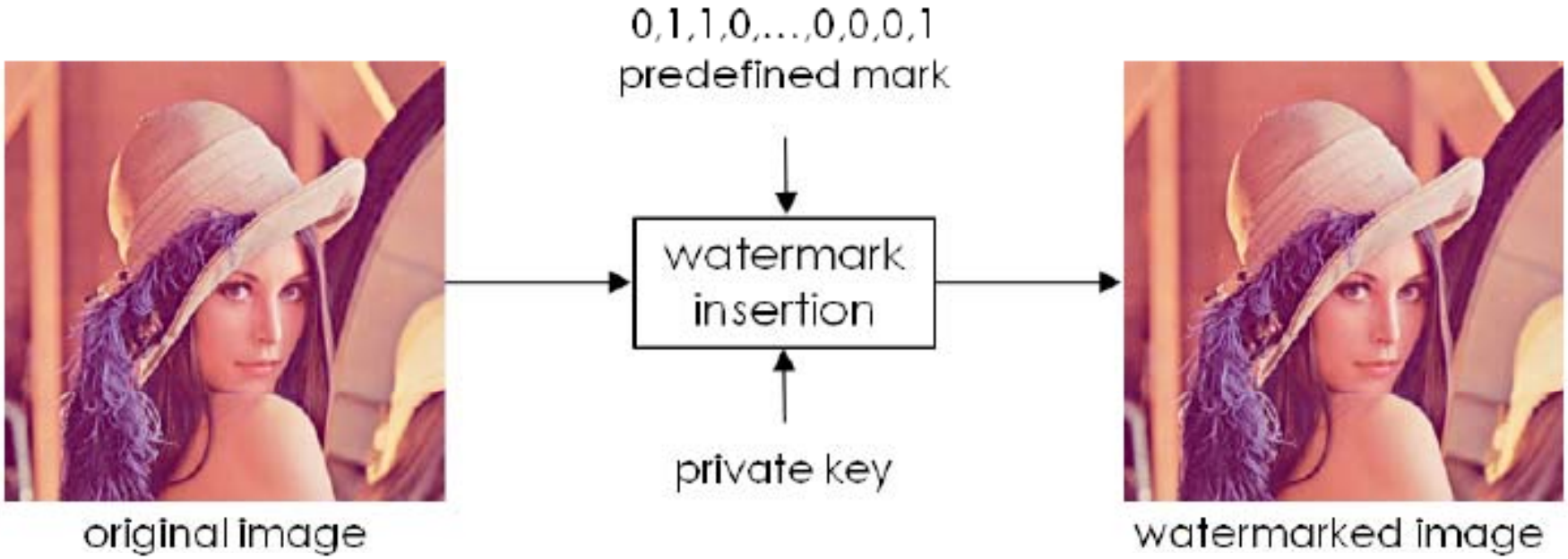
(3) **定位能力**：系统应能准确定位恶意篡改发生的具体位置，如判断图像的哪块区域被恶意改动过；

(4) **恢复能力**：针对所检测定位出的篡改区域，有时可能要求系统具备进一步恢复其原先真实内容的能力；这种恢复可以是粗略的或高精度的，视具体应用要求而定。

内容认证：脆弱水印

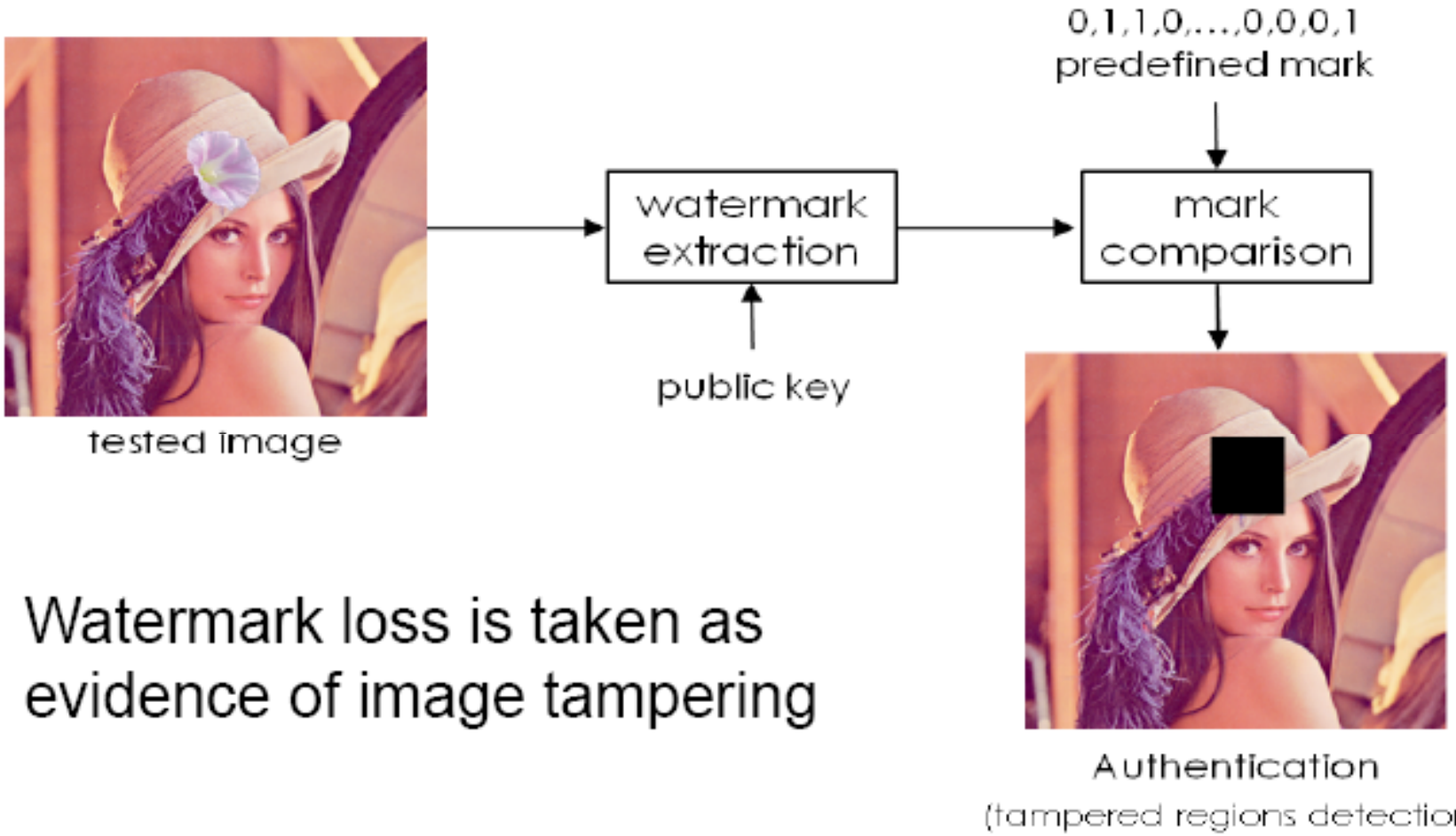
Authentication via fragile watermarking

- A fragile watermark is lost as soon as the image is modified



内容认证：脆弱水印

Authentication via fragile watermarking

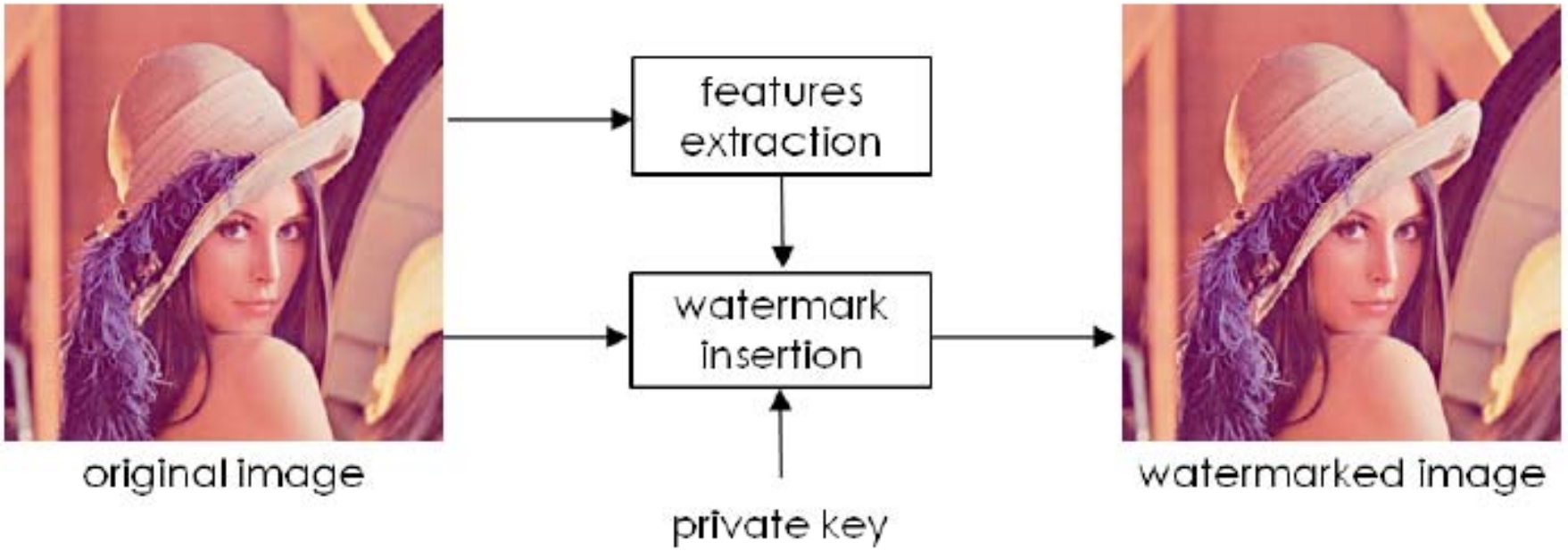


- Watermark loss is taken as evidence of image tampering

内容认证：半脆弱水印

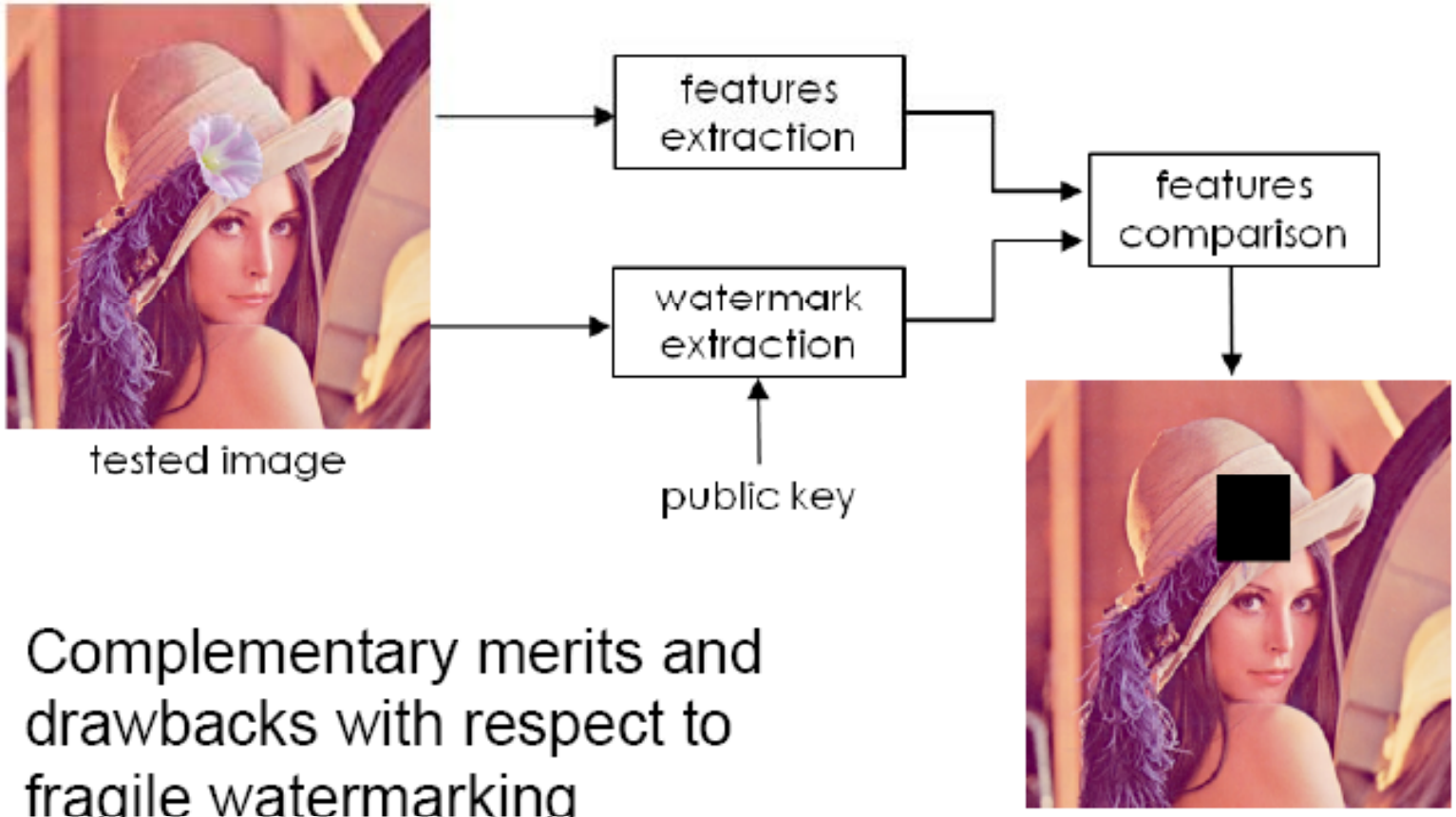
Authentication via robust watermarking

- With robust watermarking a summary of the image is inserted within the image itself



内容认证：半脆弱水印

Authentication via robust watermarking

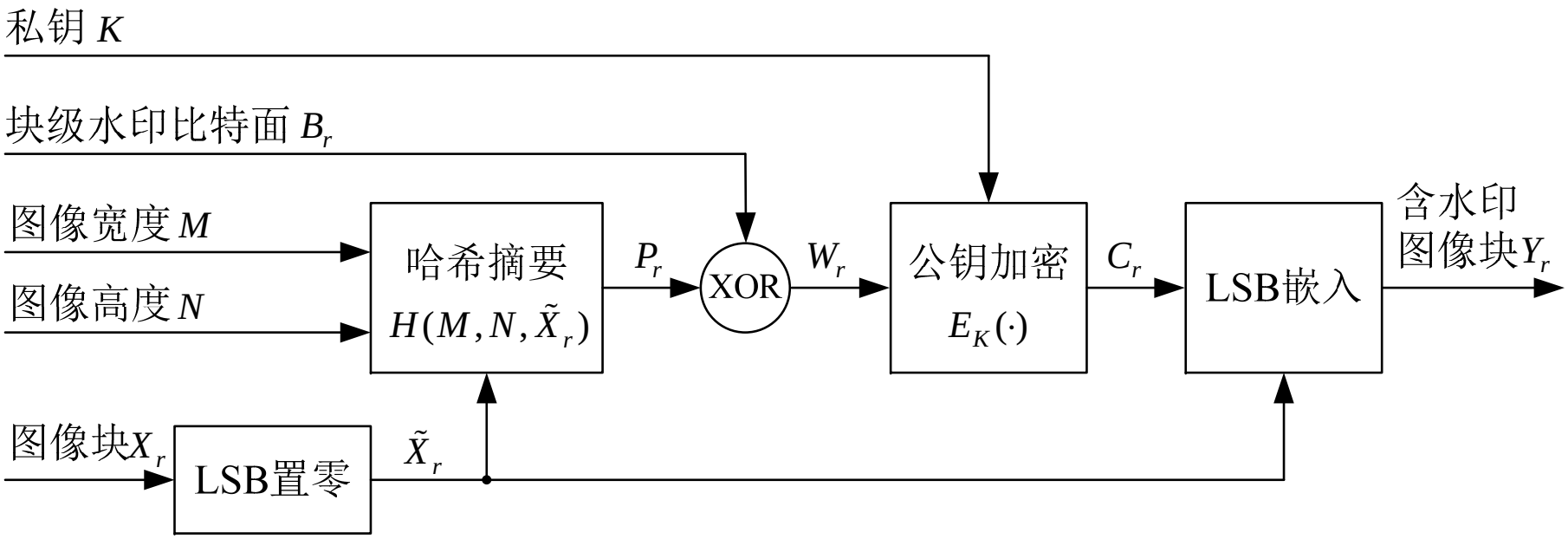


- Complementary merits and drawbacks with respect to fragile watermarking

典型算法⑥

Wong脆弱水印算法

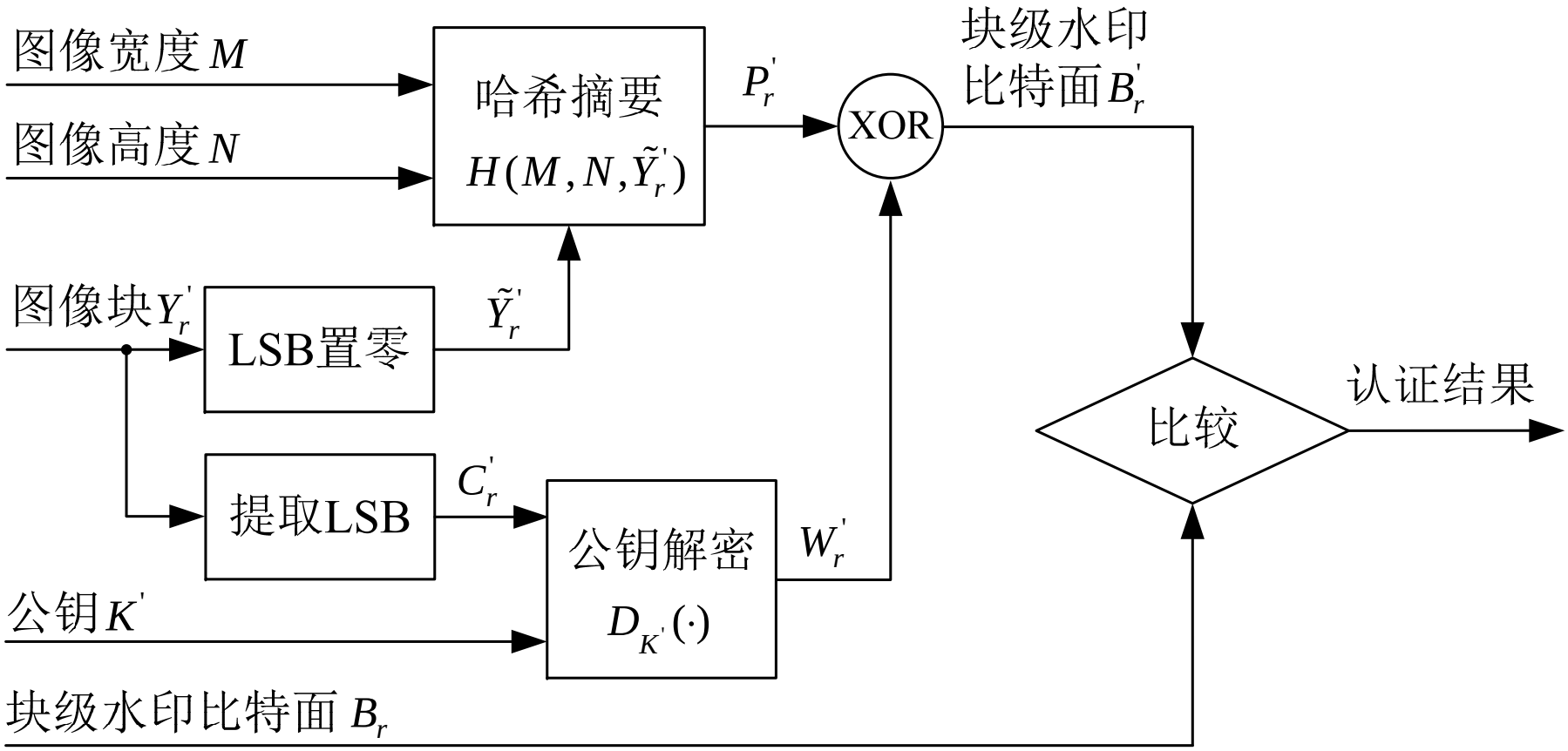
水印嵌入流程图



[1] A Public Key Watermark for Image Verification and Authentication_ICIP1998 (cite500+)

[2] Secret and Public Key Image Watermarking Schemes for Image Authentication and Ownership Verification_TIP2001 (cite500+)

水印提取/验证流程图


图像宽度 M
图像高度 N
哈希摘要
$H(M,N,\tilde{Y}_r')$
P_r'
XOR
块级水印
比特面 B_r'
图像块 Y_r'
LSB置零
$\tilde{Y}_r'$
比较
认证结果
提取LSB
C_r'
公钥解密
$D_{K'}(\cdot)$
W_r'
公钥 K'
块级水印比特面 B_r

水印提取/验证流程图

需要指出的是，Wong 脆弱水印方案利用了异或运算的如下特性：如果$c = a \oplus b$，那么应有$a = b \oplus c$和$b = a \oplus c$。不难看出，对于待测图像未经历过任何改动的情形，即$Y_r^{'} = Y_r$，在验证阶段计算$B_r^{'} = P_r^{'} \oplus W_r^{'}$时，由于$W_r^{'} = W_r$，$P_r^{'} = P_r$，因此$B_r^{'} = P_r \oplus W_r = B_r$，故该算法能做出正确的认证结果。如果待测图像经历过改动，则无法保证$W_r^{'} = W_r$，$P_r^{'} = P_r$同时成立，从而使得$B_r^{'} \neq B_r$，算法也能做出正确判决。

典型算法⑦

一种可抵抗JPEG压缩的半脆弱水印算法

C.-Y. Lin and S.-F. Chang. Semi-fragile watermarking for authenticating JPEG visual content. Proc. of SPIE 3971, Security & Watermarking of Multimedia Contents II, pp.140-151, 2000

3 数字水印技术

- 框架
- 分类
- 评价指标
- 攻击方法
- 版权保护水印
- 内容认证水印
- 可逆水印

可逆水印技术

可逆数字水印（**Reversible watermark**）技术属于数字水印技术的一个分支，目前大多数数字水印的方法在提取出所嵌入的秘密信息后，原宿主信息不能无损恢复，属于有损数字水印技术。但是在一些要求较高的场合，如医学诊断、军事图像、遥感图像处理及法律认证及证据等领域，则往往需要精确地恢复原载体。

可逆数字水印评价指标

（1）畸变程度：评价秘密信息嵌入后的畸变程度有两种方式，一种是主观评价法，另一个是客观评价法。

主观评价法将人对载体的感觉分为几个等级，该方法易受到评价者主观因素的影响；常用的客观评价方法主要有均方差和峰值信噪比。

可逆数字水印评价指标

（2）嵌入率：嵌入量就是嵌入到载体中的秘密信息的比特数。嵌入量可以分为实际嵌入量和有效嵌入量。实际嵌入量即数字水印方法本身可嵌入的比特数，而有效嵌入量是指实际可隐藏的嵌入量再减去附加信息的信息量。

常用的有效嵌入量的指标是嵌入率 **ER**，其表达式为：

$$ER = \frac{Num_{\text{sec}} - Num_{extra}}{Num_{feature}} (bit / \ feature, bpf)$$

可逆数字水印技术分类

（1）基于无损压缩的可逆数字水印方法

（2）基于差值扩展的可逆数字水印方法

（3）基于直方图修改的可逆数字水印方法（要求掌握）

典型算法⑧

基于直方图修改的可逆水印

基于直方图修改的可逆水印技术

以图像为例，水印嵌入的步骤如下：

（1）首先计算图像的直方图，并找到其中的零点，即图像中没有任何灰度值的像素点记为 z；然后找到直方图像素点最多的灰度值的峰值点，记为 p。为了方便叙述，不妨假设 $p < z$ 。

（2）由上到下、由左到右扫描图像中的各个像素点，各个像素点的灰度值用 v_{ij} 表示，当 $v_{ij} < p$ 或 $v_{ij} > z$ 时，像素点的值保持不变，即：$v'_{ij} = v_{ij}$；否则，像素点的灰度值加1，即：$v'_{ij} = v_{ij} + 1$。

基于直方图修改的可逆水印技术

（3）图像中灰度值等于峰值点的像素点，为可嵌入秘密信息的点，将秘密信息转化为二进制流，用 S_k 表示。顺序嵌入信息得 $v'_{ij} = v_{ij} + s_k$ 。

（4）得到的由灰度值组成的图像就是嵌入秘密信息后的图像。同时 p、z 以密钥的形式保存。

基于直方图修改的可逆水印技术

秘密信息提取和原始图像的恢复过程如下：

（1）读取密钥，得到 p、z 的值。

（2）逐行扫描图像，当 $v_{ij} = p$ 时，说明该点为隐藏信息点，提取信息0并保持该点灰度值不变；当 $v_{ij} = p+1$ 时，该点也为隐藏信息点，提取信息1并使该像素点值减1。

（3） 当 $v_{ij} < p$ 或 $v_{ij} > z$ 时，像素点的值保持不变；当 $p-1 < v_{ij} < z$ 时，像素点的灰度值减1。

（4）得到由灰度值组成的新图像就是提取秘密信息后的恢复出来的载体图像。

基于直方图修改的可逆水印技术

基于直方图修改的可逆数字水印，具有以下优点：

（1）产生较少的畸变，具有较高的峰值信息比。

（2）对于使用该方法直接在空域中应用不会产生数据溢出问题。

（3）对于某些含有大量相同背景的图像具有较高的嵌入率。这对于一些数字医学图像具有很好的应用效果。

基于直方图修改的可逆水印技术

缺点：由于图像直方图的直接使用，因此其嵌入率不稳定，对于一般图像嵌入率较低。

改进方法：

（1）选择无穷远处为其差值直方图的零值点。

（2）可选择两个或多个峰值进行可逆信息隐藏。

（3）可以进行多层隐藏。

主要内容

1. 信息隐藏基本理论
2. 空域/变换域信息隐藏技术
3. 数字水印
4. 应用与发展

4 信息隐藏的应用与发展

- 信息隐藏技术的应用与发展方向

  （1）数字知识的产权保护 （鲁棒水印）

  （2）数据完整性鉴定、内容认证 （脆弱/半脆弱水印）

  （3）数据保密

  （4）资料不可抵赖性的确认

4 信息隐藏的应用与发展

信息隐藏技术为信息安全提供了一种新的思路，为我们研究信息安全提供了一个新的方向。但总的来说，信息隐藏技术尚没有发展到可普及实用的阶段，使用密码加密仍是网络信息传输的主要安全手段。现在还存在大量的实际问题亟待解决，如信息隐藏的容量问题，如何建立不可感知性的数学度量模型，信息隐藏的容量上界如何计算等。

4 信息隐藏的应用与发展

- 数字水印技术的应用和发展方向

（1）很多研究在试图使水印处理技术与编码算法统一起来。这样可以使水印对该编码算法具有鲁棒性，尽量减少无意的水印攻击。

（2）许多组织力图建立国际统一标准的水印算法。

（3）很多公司和研究机构着力将水印处理技术商业化并应用于其他领域如军事和国防领域，用于传送秘密的军事命令、验证军事命令、信息的真实可靠性。

要求掌握的核心内容

- 信息隐藏的概念、技术分类、性能指标
- 理解并掌握典型信息隐藏算法（包括LSB系列，Patchwork，基于扩频(SS)，QIM，Jsteg 系列，Wong算法，基于直方图修改的可逆水印等）
- 数字水印概念、技术分类、性能指标
- 信息隐藏与数字水印的区别与联系
- 版权保护水印技术和内容认证水印技术的基本原理与应用方法

思考题

- 信息隐藏与数字水印的区别与联系。
- 例举常见的信息隐藏技术。
- 简单描述数字水印的组成框架。
- 数字水印有哪些攻击，分别有何应对策略。
- 简述各典型数字水印算法的基本原理与、主要步骤、优缺点以及改进方法。

附：图像局部隐藏的应用

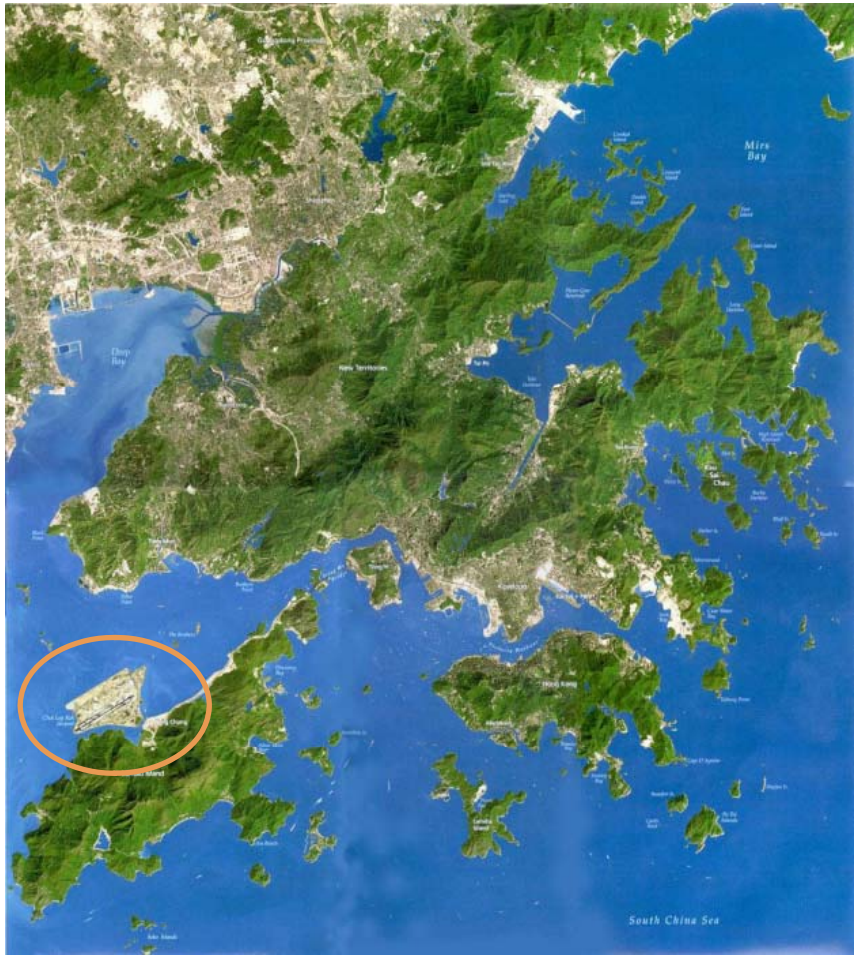
- 图像局部隐藏工具
- 研究目标：将图像中的局部关键信息剪切后自动填补，并嵌入被剪切的信息。合法用户可以使用约定的密钥获得原始图像，其他接收者则无法获得完整的原始图像

图像局部隐藏

局部图像恢复

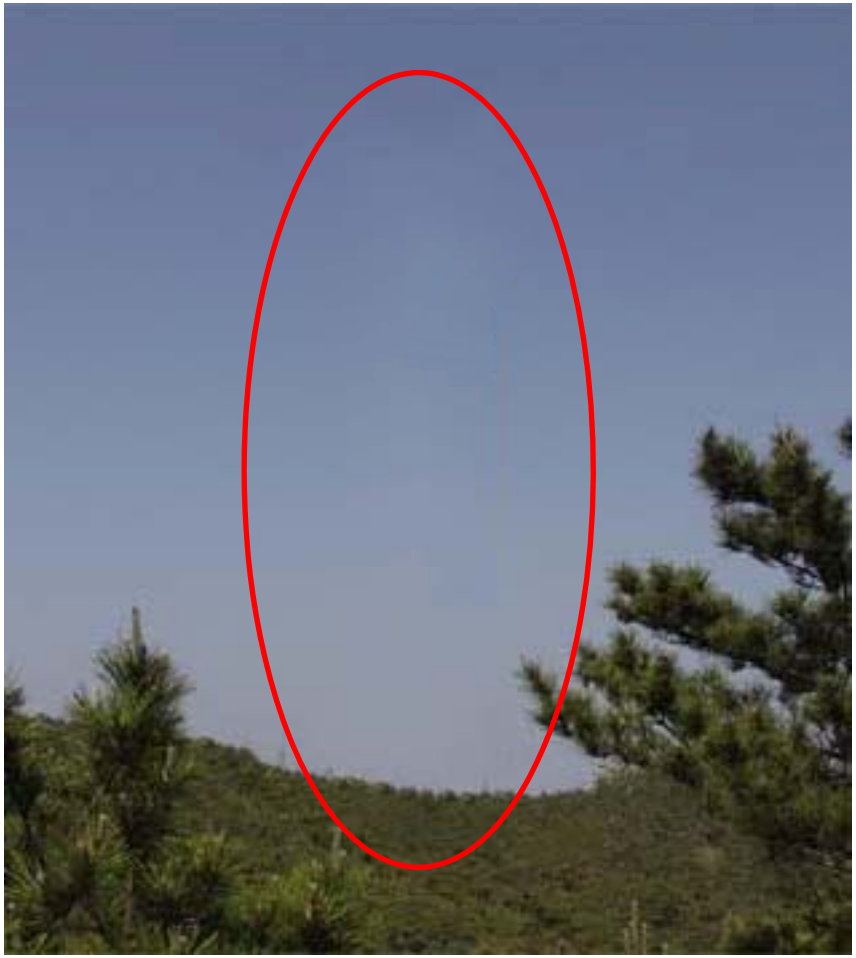
局部图像隐藏

图像局部隐藏



香港地形图

图像局部隐藏



山顶雷达站