Digital Watermarking and Steganography

by Ingemar Cox, Matthew Miller, Jeffrey Bloom, Jessica Fridrich, Ton Kalker

Chapter 3. Models of Watermarking

Lecturer: Jin HUANG

Overview

Several conceptual models of watermarking

- View of communications 通信
- View of geometry 几何

Correlation-based watermarking

- How to measure “it is THE message”.

3.2 Communications

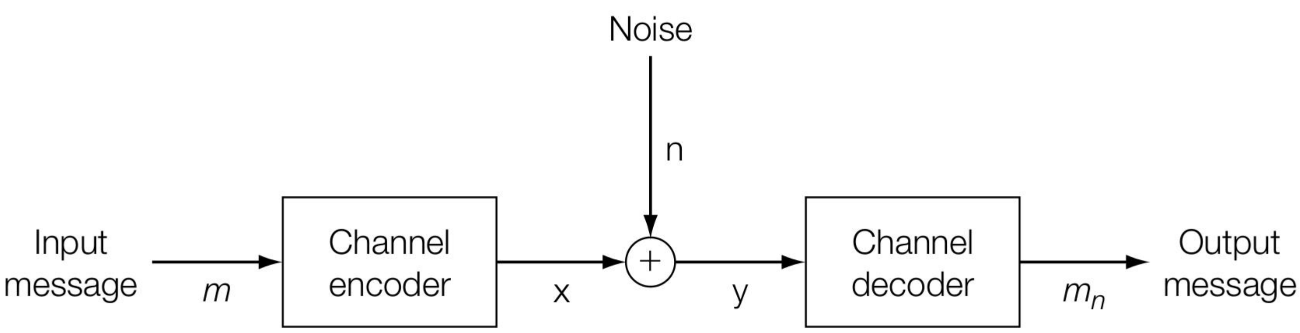
Components of Communications Systems

无法传输 编码 解码后

message → $\boxed{x \to y}$ → m_n

信道中

- $\mathbf{x}$ is signal that can be transmitted over the channel, but m is not.
  - Source coder: draw symbols in some alphabet.
  - Modulator: converts a sequence of symbols into a physical signal.
- Transmission in channel add noise $\mathbf{n}$.



Classes of Transmission Channels

According to the type of noise function

- Additive noise: $\mathbf{y} = \mathbf{x} + \mathbf{n}$.
- Fading channel: $\mathbf{y} = \nu[t]\mathbf{x} + \mathbf{n}, 0 \leq \nu[t] \leq 1$.
- ...

eg. 声音、车方衰减

Secure Transmission 1

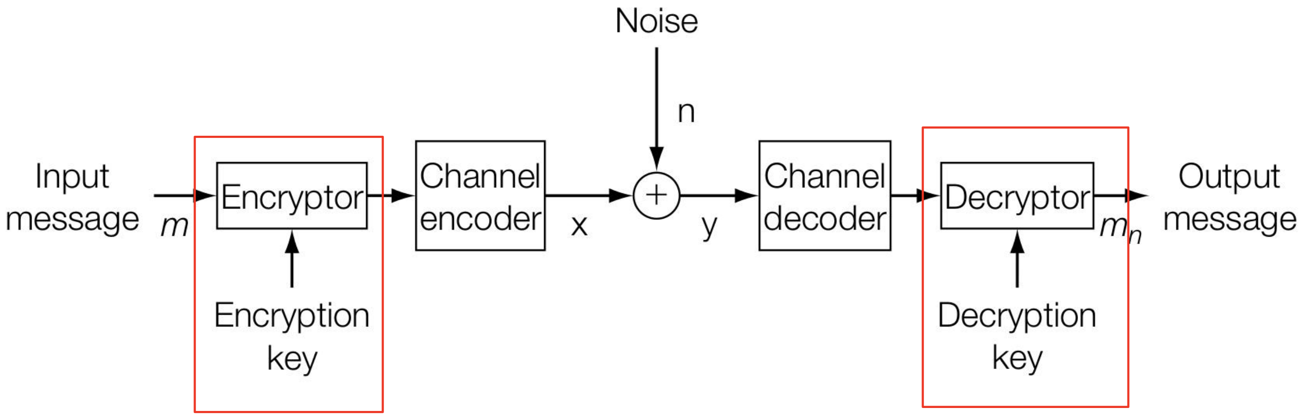
Security against both passive and active adversaries

- Passive: Aims at the message.
  - Monitors the transmission channel and attempts to illicitly read the message. 监听信息
- Active: Aims at the transmission.
  - Disable the communications or transmit fake/unauthorized messages. 阻断传输

Secure Transmission 1

Message layer: cryptography.

- Prevent unauthorized reading.
- Prevent unauthorized writing.

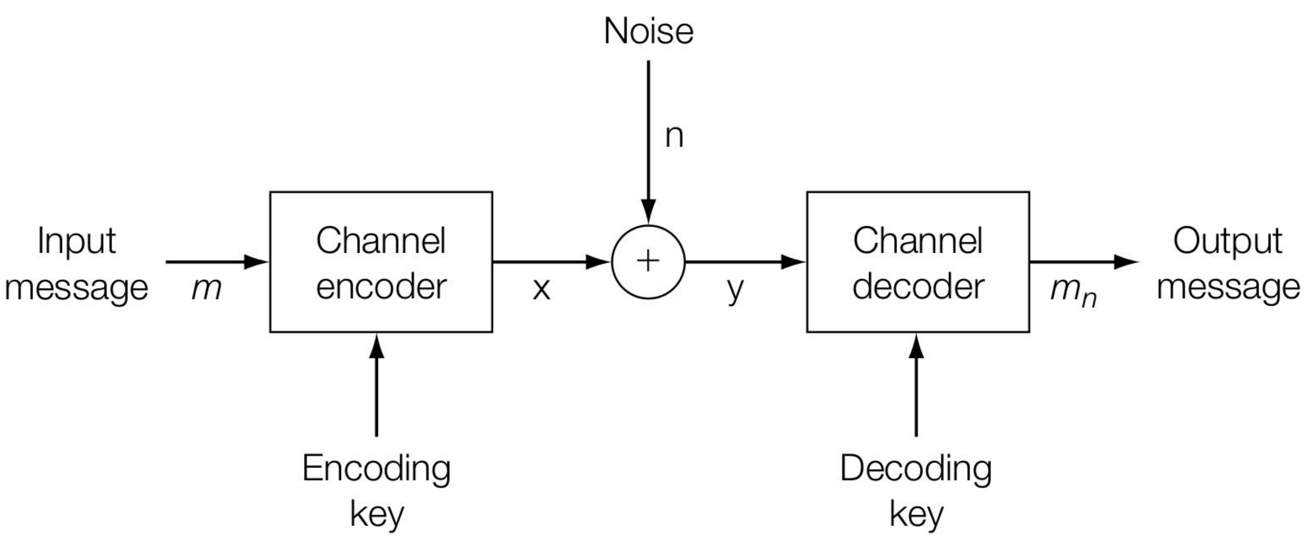


Secure Transmission 2

光谱

Transport layer: spread spectrum communication.

- Spreads the signal across a wider bandwidth according to a secret key.
  - Frequency hopping. 每次交流不同频率 大功率无线电信号压制
  - Cannot monitor the transmission.
  - Huge cost/power to jam the transmission.



3.3 Communication-Based Models of Watermarking

Models

Deliver the message from the embedder to decoder.

原文件 水印

- Not suitable for authentication system. (不包括认证) 表示不能修改

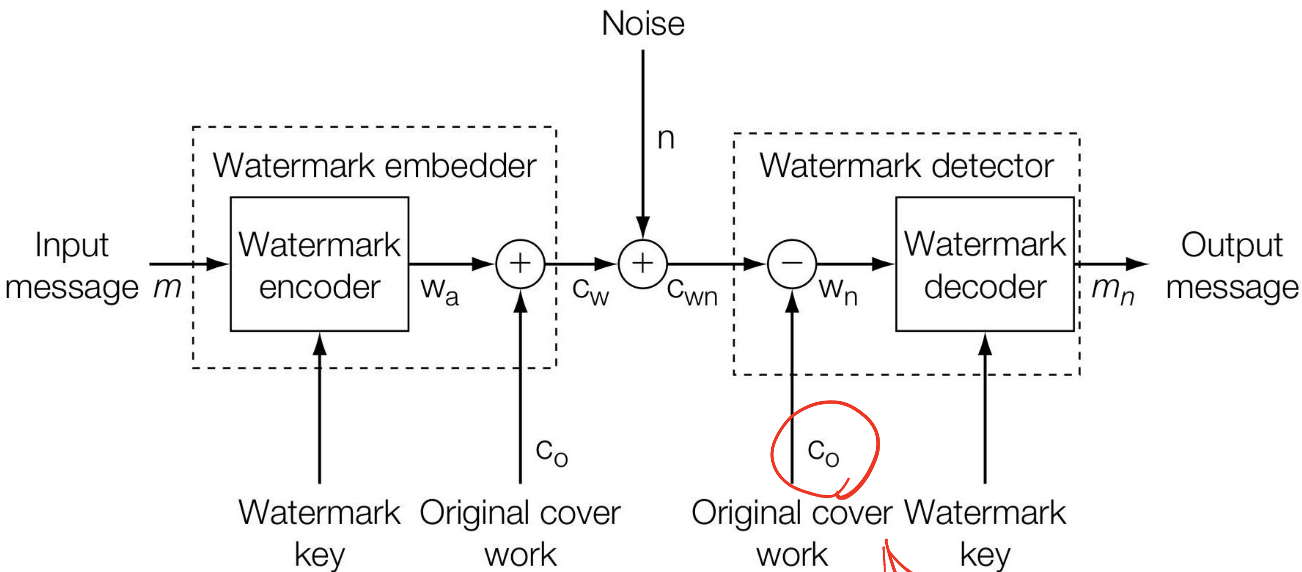
$$\mathbf{c}_{wn} = \mathbf{c}_o + \mathbf{w}_a + \mathbf{n}$$

信道中加入的

How to use the cover work.

- As noise.
- As side information. 辅助W传输
- The second message.

As Noise 1

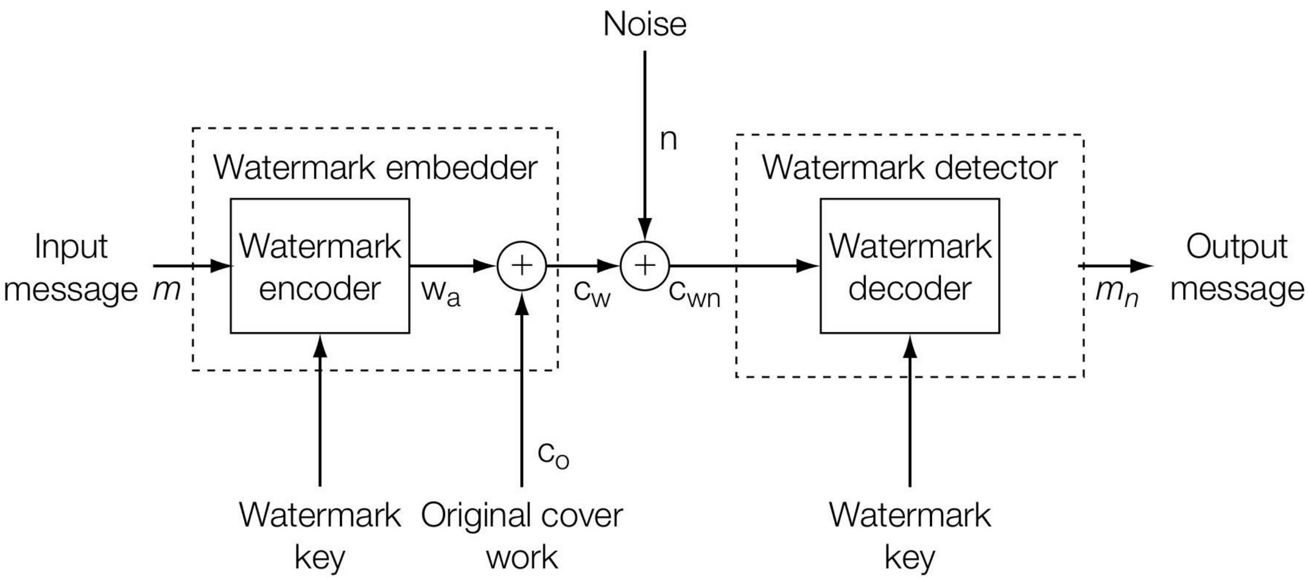


Informed Detector

需要另一种途径把c_o传输

To cancel out effect of $\mathbf{c}_o$, the whole $\mathbf{c}_o$ is not always required.

As Noise 2



Blind Detector

$\mathbf{w}_a$ is corrupted by both $\mathbf{c}_o$ and $\mathbf{n}$.

Blind Embedding (E_BLIND)

One bit only message $m \in 0, 1$:

- A **reference pattern** (key) $\mathbf{w}_r$. 和C_0一样大
- Encoding into to **message pattern**:

$$\mathbf{w}_m = (2m - 1)\mathbf{w}_r.$$

防止加上去会超过255
小于0.

- Modulate to **added pattern**: $\mathbf{w}_a = \alpha \mathbf{w}_m$.
- Embedding: $\mathbf{c}_w = \mathbf{c}_o + \mathbf{w}_a$.

Linear Correlation Decoder (D_LC)

After transmission $\mathbf{c} = \mathbf{c}_w + \mathbf{n}$.

Detection:

- Goal: How $\mathbf{c}$ is correlated to $\mathbf{w}_r$?
- Linear Correlation (scaled dot product):

$$z_{lc}(\mathbf{c}, \mathbf{w}_r) = \frac{1}{N}\mathbf{c} \cdot \mathbf{w}_r, \quad \mathbf{c} \in \mathbb{R}^N.$$

图像大小

  - Larger $|z_{lc}|$ means higher correlation.
  - An imperfect measurement (will show later).

相关性.

Why Dot Product?

Start from the usual distance definition:

$$\sum_i (\mathbf{a}_i - \mathbf{b}_i)^2 = \|\mathbf{a} - \mathbf{b}\|^2$$

Why Dot Product?

Start from the usual distance definition:

$$
\begin{aligned}
\sum_i (\mathbf{a}_i - \mathbf{b}_i)^2 &= \|\mathbf{a} - \mathbf{b}\|^2 \\
&= (\mathbf{a} - \mathbf{b})^T (\mathbf{a} - \mathbf{b})
\end{aligned}
$$

Why Dot Product?

Start from the usual distance definition:

$$
\begin{aligned}
\sum_i (\mathbf{a}_i - \mathbf{b}_i)^2 &= \|\mathbf{a} - \mathbf{b}\|^2 \\
&= (\mathbf{a} - \mathbf{b})^T (\mathbf{a} - \mathbf{b}) \\
&= \mathbf{a}^T \mathbf{a} - 2\mathbf{a}^T \mathbf{b} + \mathbf{b}^T \mathbf{b}
\end{aligned}
$$

Why Dot Product?

Start from the usual distance definition:

$$
\begin{aligned}
\sum_i (\mathbf{a}_i - \mathbf{b}_i)^2 &= \|\mathbf{a} - \mathbf{b}\|^2 \\
&= (\mathbf{a} - \mathbf{b})^T (\mathbf{a} - \mathbf{b}) \\
&= \mathbf{a}^T \mathbf{a} - 2\mathbf{a}^T \mathbf{b} + \mathbf{b}^T \mathbf{b} \\
&= (\|\mathbf{a}\|^2 + \|\mathbf{b}\|^2) - 2\mathbf{a} \cdot \mathbf{b}.
\end{aligned}
$$

间隔近 → 点积大

z_{lc}

Assuming $\mathbf{c}_o, \mathbf{n}$ are from Gaussian distributions:

$$
\begin{aligned}
z_{lc} &= \frac{1}{N}(\mathbf{c}_o + \mathbf{w}_a + \mathbf{n}) \cdot \mathbf{w}_r \\
&= \frac{1}{N}\left(\mathbf{w}_a \cdot \mathbf{w}_r + \underline{(\mathbf{c}_o + \mathbf{n}) \cdot \mathbf{w}_r}\right) \\
&= \frac{1}{N}(\mathbf{w}_a \cdot \mathbf{w}_r) + \varepsilon \\
&= \frac{1}{N}(\alpha(2m-1)\mathbf{w}_r \cdot \mathbf{w}_r) + \varepsilon \\
&= (2m-1)\left(\alpha \frac{\|\mathbf{w}_r\|^2}{N}\right) + \varepsilon.
\end{aligned}
$$

高维随机向量几乎都互相垂直

与均值和方差有关

m_n

Decoder outputs

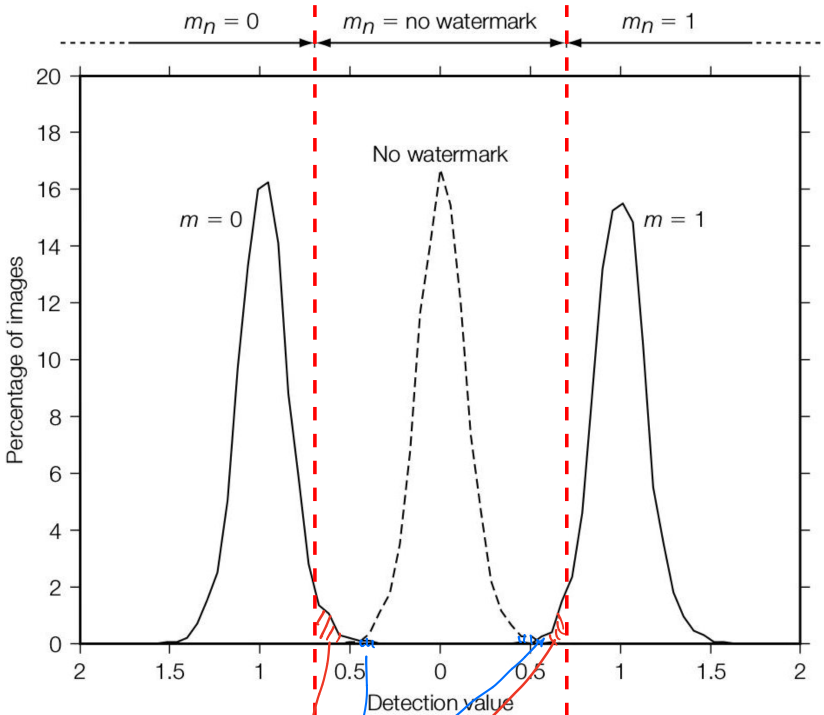
$$m_n = \begin{cases} 1 & z_{lc} > \tau_{lc} \\ \text{no} & -\tau_{lc} \leq z_{lc} \leq \tau_{lc} \\ 0 & z_{lc} < -\tau_{lc}. \end{cases}$$

- $\alpha = 0 \Leftrightarrow$ no.
- τ_{lc} is important.

Testing Parameters

- Unit variance: $\sigma^2_{\mathbf{w}_r} = \|\mathbf{w}_r - \mu_{\mathbf{w}_r}\|^2/N = 1$.
  - $\mu_{\mathbf{x}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{x}[i]$.
  - $\sigma^2_{\mathbf{x}} = \mu_{(\mathbf{x}[i]-\mu_{\mathbf{x}})^2} = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{x}[i] - \mu_x)^2$.
- 2000 images for $\mathbf{c}_o$, 6000 images as $\mathbf{c}_w$.
  - 2000: $\alpha = 0$, no watermark.
  - 2000: $\alpha = 1, m = 1$.
  - 2000: $\alpha = 1, m = 0$.
- $\tau_{lc} = 0.7$. 为什么是0.7
  - False positive probability $P_{fp} \approx 10^{-4}$.
  - In Chapter 7.

Performance



考试：

如何看图.

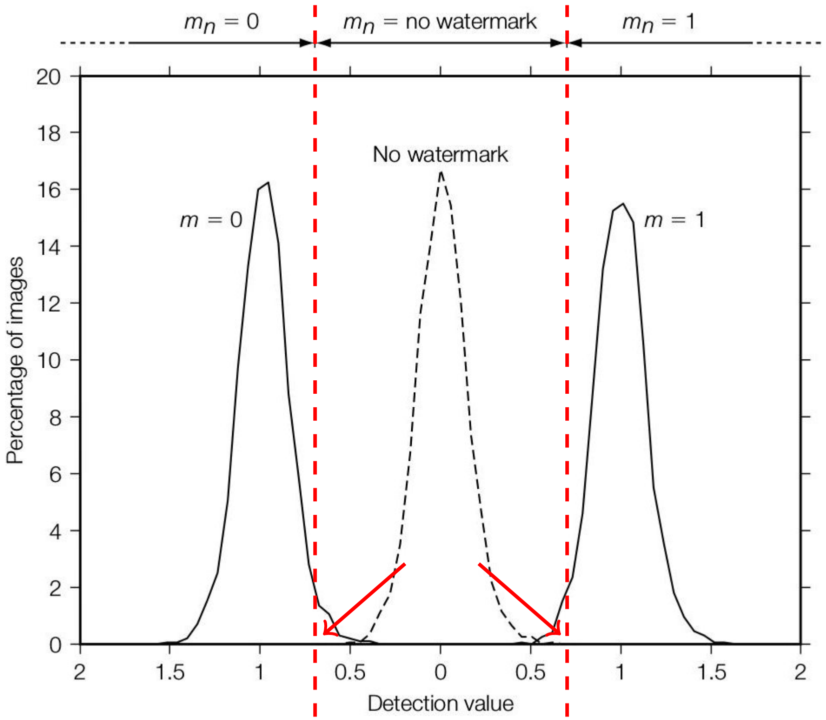
eg. 画 fpr 对应

图中哪一块

effectiveness.

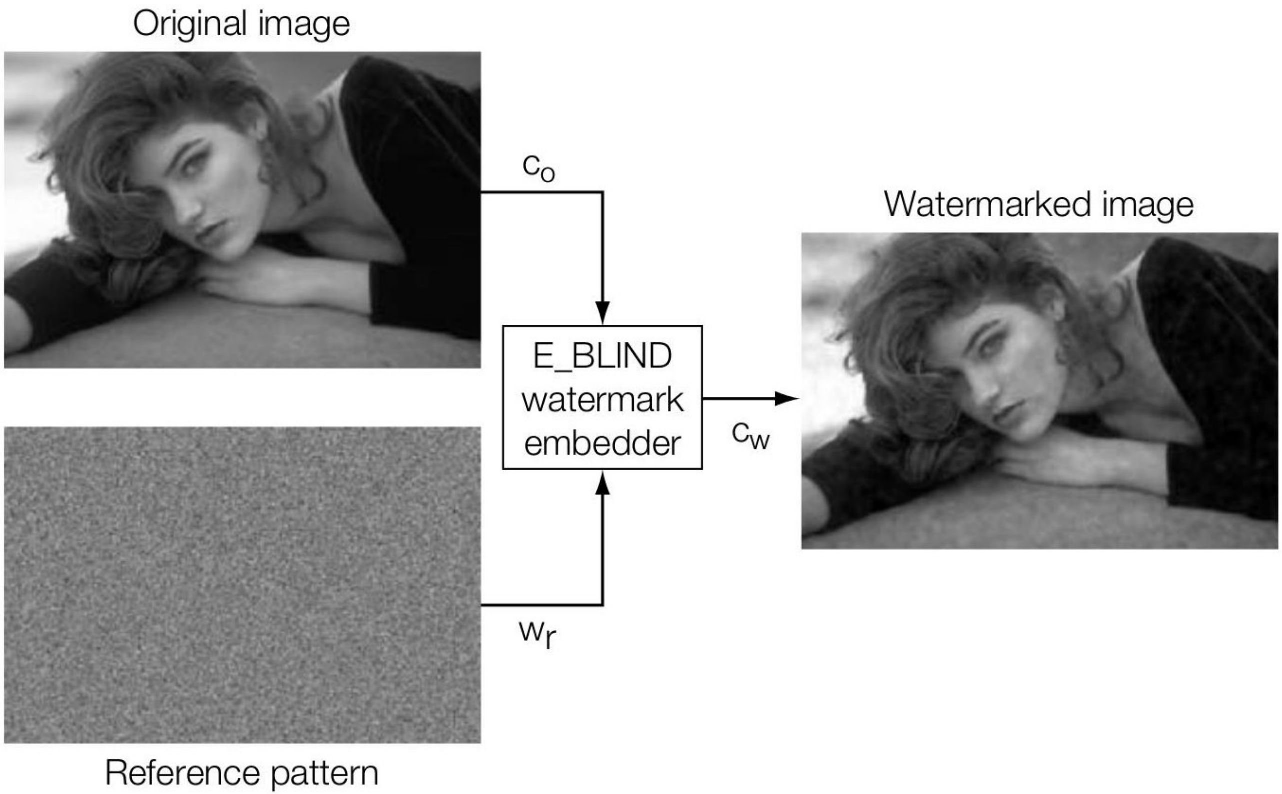
- False positive rate: 0.01%.
- Effectiveness: $1 - (57 + 41)/4000 \approx 98\%$.

Performance



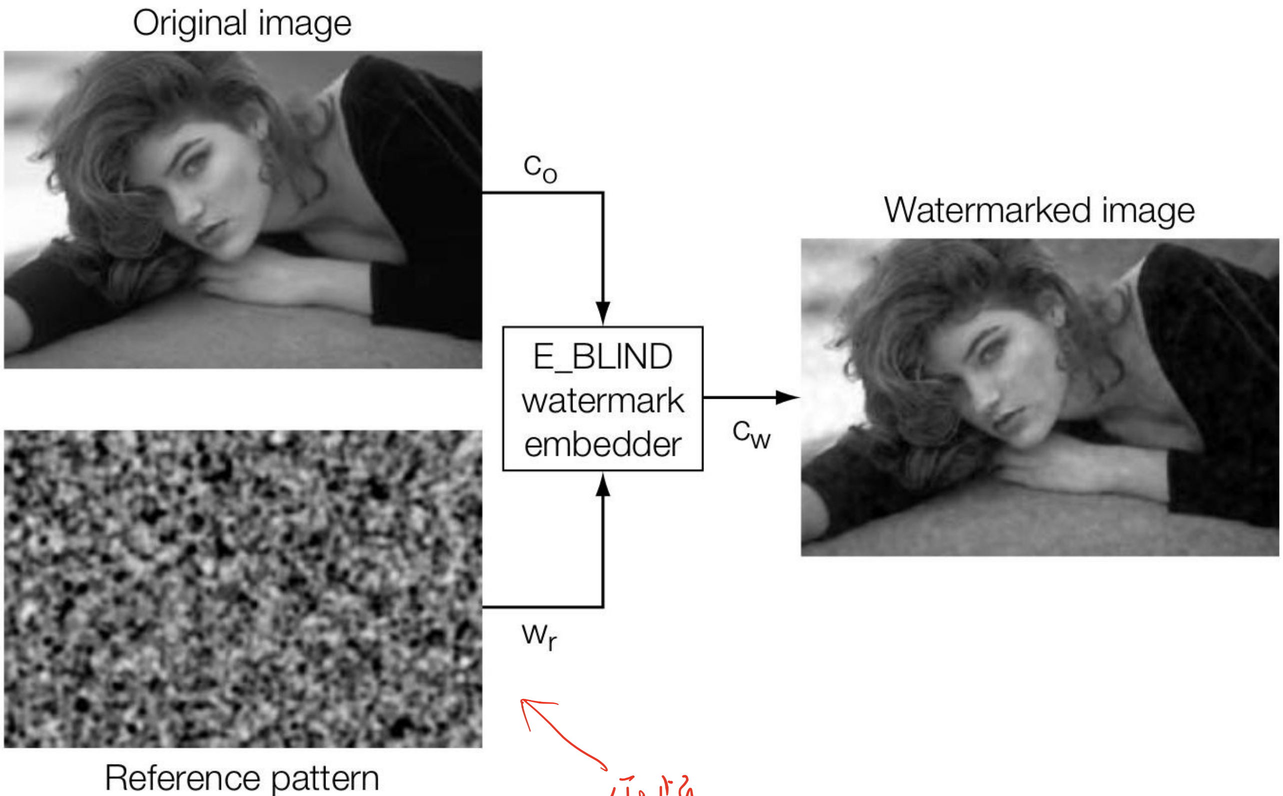
- False positive rate: 0.01%.
- Effectiveness: $1 - (57 + 41)/4000 \approx 98\%$.

High Frequency Reference



Pseudo-random number for each pixel.

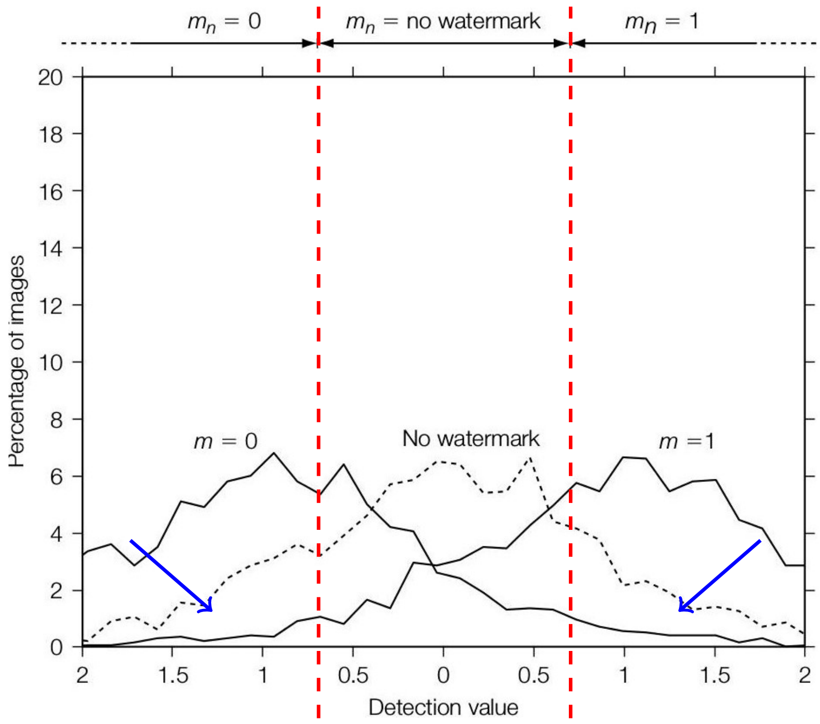
Low Frequency Reference



低频

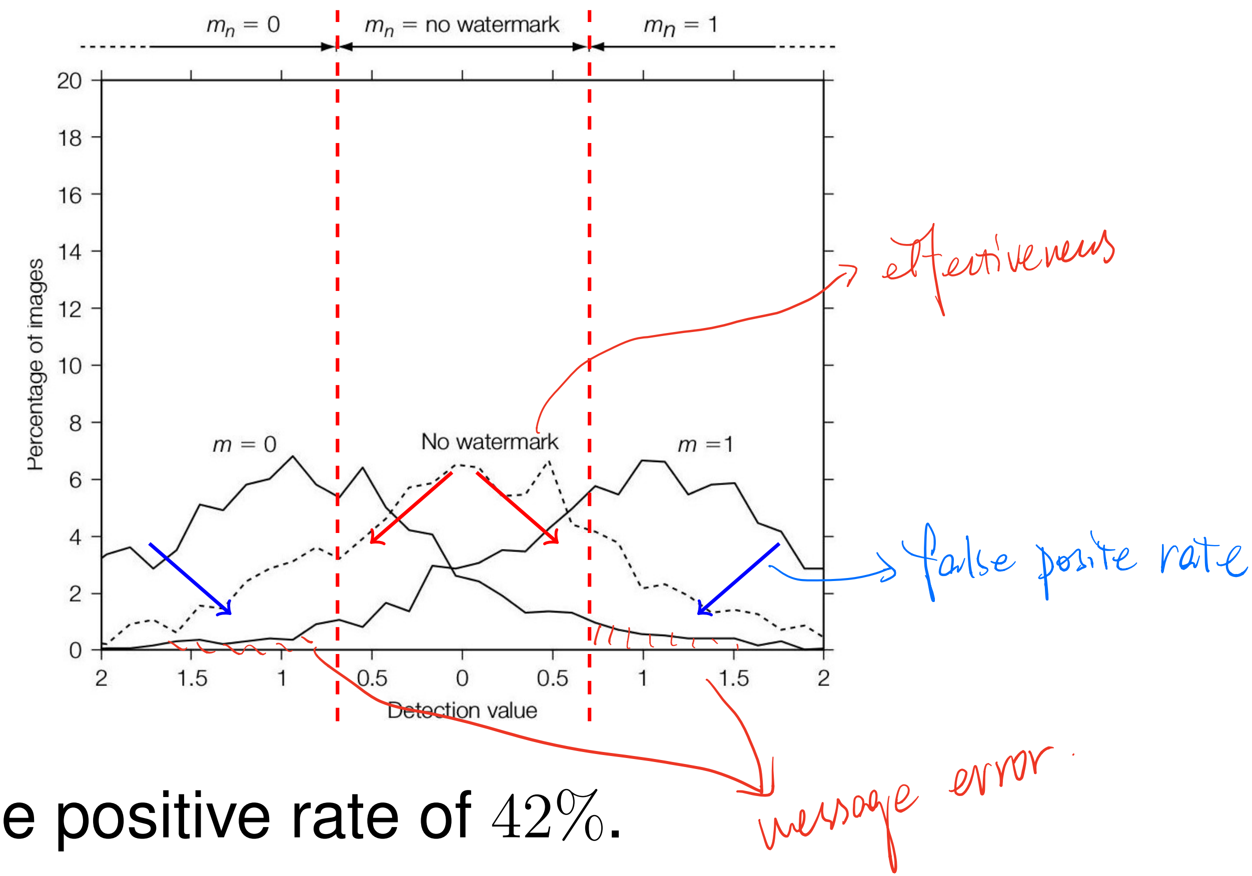
Applying a low-pass filter. Worse fidelity.

Worse Performance



- False positive rate of 42%.
- Effectiveness: 68%.

Worse Performance



- False positive rate of 42%.
- Effectiveness: 68%.

Reason

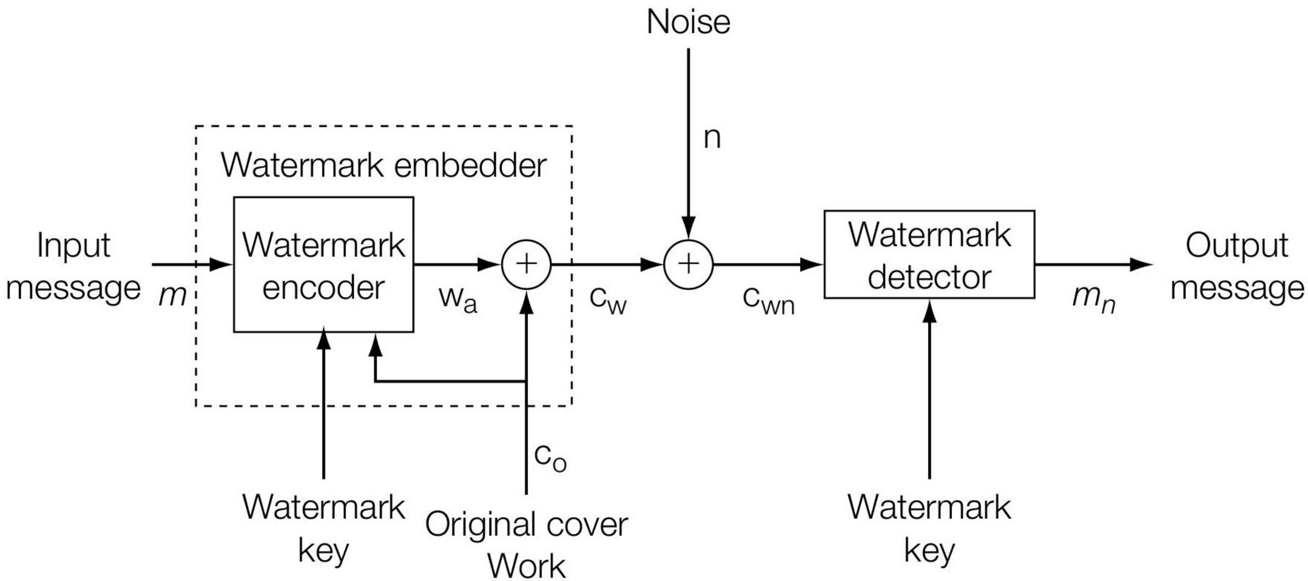
图像和噪音都是低频（分辨率降低）

$W_r \cdot C_0$ 变大

ε is large:

- High inherent correlations between the images and the reference pattern.
- Images tend to have more energy in the low frequencies than in the high.

Help from $\mathbf{c}_o$



$\mathbf{c}_o$ is part of the noise.

- We know it, and use it for
  - 100% effectiveness!

Embedding with Side Information

Adaptive strength α: 寻找α.

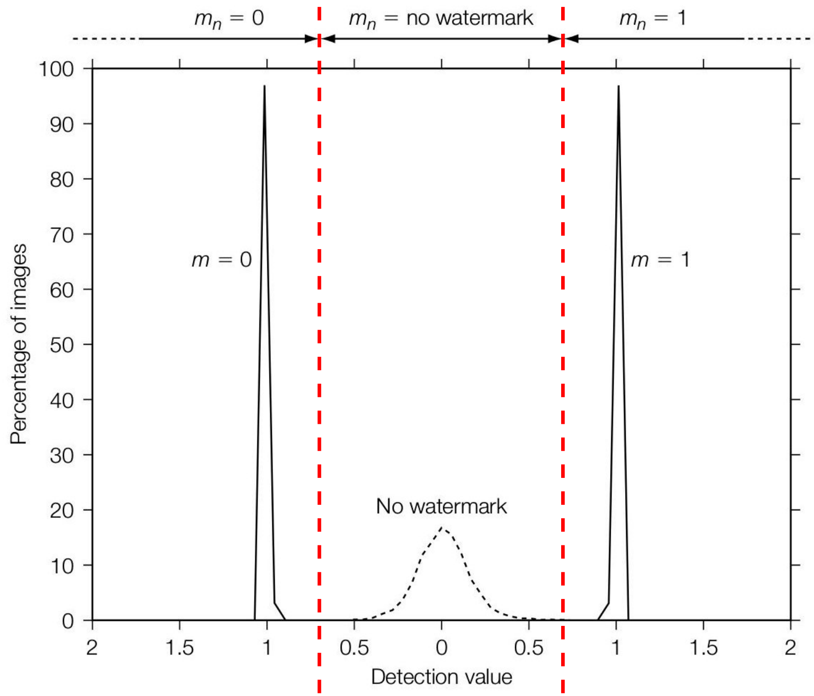
- Correlation must be large enough:

$$
\begin{aligned}
\tau_{lc} < \tau_{lc} + \beta &= z_{lc}(\mathbf{c}_w, \mathbf{w}_m) \\
&= \frac{1}{N}(\mathbf{c}_o + \alpha \mathbf{w}_m) \cdot \mathbf{w}_m. \\
\Longrightarrow \alpha &= \frac{N(\tau_{lc} + \beta) - \mathbf{c}_o \cdot \mathbf{w}_m}{\mathbf{w}_m \cdot \mathbf{w}_m}.
\end{aligned}
$$

- May sacrifice fidelity.

Performance

考点：↓虚线不会变



判断题：通过改进 encoder 来提高fps ✗.

不加水印，和偏移 有标形差

- False positive rate of 0.01%. 不变
- Effectiveness: 100%. 提高↑

Discussion

- How about directly making $\varepsilon = 0$?
  - Find an approximation $\mathbf{c}'_o$ so that

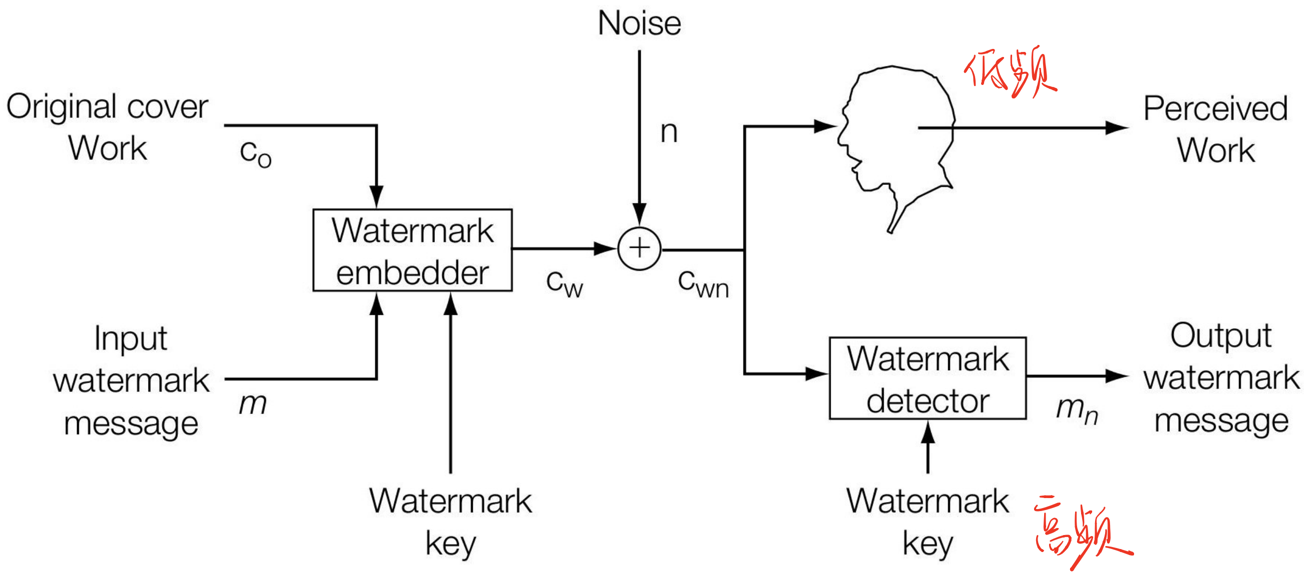
$$\mathbf{c}'_o \cdot \mathbf{w}_m = 0.$$

  - How?

$$\mathbf{c}'_o = \mathbf{c}_o - \frac{\mathbf{c}_o \cdot \mathbf{w}_m}{\mathbf{w}_m \cdot \mathbf{w}_m} \mathbf{w}_m.$$

  - Is it good?
    - Equivalent to ?
- Will false positive be zero? 不会.
  - Murphy's law: Anything that can go wrong will go wrong (Interstellar).

Multiplexed Communications 1



Multiplexed Communications 2

- In traditional communications:
  - Same method but different parameter
    - Time, frequency, or code sequence.
- In watermarking:
  - Different methods
    - Frequency division for one
    - Spread spectrum coding for the other.
- Signal-to-noise ratio (SNR)
  - Which one is the signal.

Project: System 1

- E_BLIND
- D_LC

Presentation: 7.3,7.4

- False Negative Errors
- ROC curve
  - Receiver operating characteristic curve
  - Balance of false positives and false negatives rate.