Watermarking embedding and detection

Embedding

- some embedding rule are independent of the particular host asset to be watermarked
- some other adapt the embedding strategy to the host asset (informed watermarking)

Additive watermarking

$$f_{w,i} = f_i + \gamma\, w_i$$

f_i i-th component of the original feature vector

w_i i-th sample of the watermarking signal

γ a scaling factor controlling the watermark strength

$f_{w,i}$ i-th component of the watermarked feature vector

Embedding

- simple
- watermark strength can be adapted to the local characteristics

$$f_{w,i} = f_i + \gamma_i w_i$$

or

$$f_{w,i} = f_i + \gamma_i (f_i) w_i$$

- under the assumption that the host features follows a Gaussian distribution and that attacks are limited to the addition of white Gaussian noise (AWGN), correlation-based decoding is optimum

⬇

- overall error probability given a false detection rate is minimized

Embedding

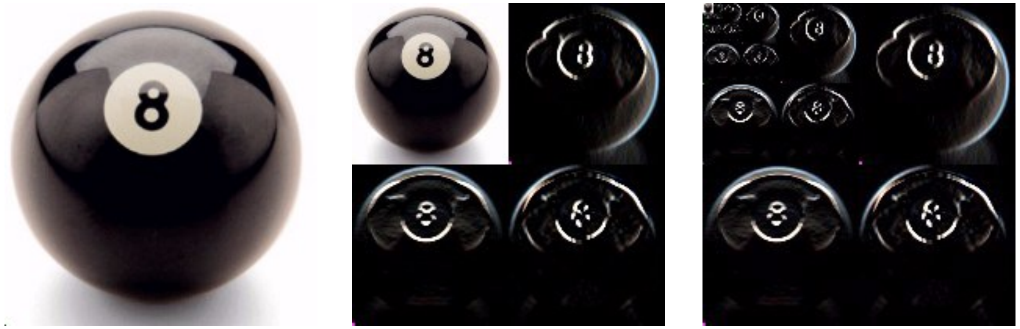
Example in “Exploiting DCT masking effect to improve the perceptual quality of data hiding” SPIE Electronic Imaging 2009

- DCT domain
- Increasing the perceptual quality of the watermarked images by exploiting the masking effect of the DCT coefficients: strength of the embedded data adaptive
- Can be exploited to i) increase watermark energy thus improving robustness, or ii) improve the perceived image quality keeping robustness constant
- 2 bits inserted into each DCT block (positions (1,3),(3,5)) using the Contrast Sensitivity Function as modulating strength

Embedding

Example in “Improved wavelet-based watermarking through pixel-wise masking” by Barni et al. IEEE Transactions on image processing, vol. 10, no. 5, May 2001

- DWT domain



- Given the watermark that has to be embedded into the wavelet coefficients it defines and exploits the just noticeable threshold of modification that each coefficient can sustain without degrading the visual quality of the image (following three perceptual rules)
- Detection based on Newman-Pearson theorem

Embedding

Example in “Improved wavelet-based watermarking through pixel-wise masking” by Barni et al. IEEE Transactions on image processing, vol. 10, no. 5, May 2001

- Mark: binary pseudorandom sequence
- Added to the DWT coefficients of the 3 largest detail subbands of the image
- DWT suitable to identify the image areas where a disturb can be more easily hidden
- It exploits a weighting function considering local sensitivity of the image to noise

Embedding

Example in “Improved wavelet-based watermarking through pixel-wise masking” by Barni et al. IEEE Transactions on image processing, vol. 10, no. 5, May 2001

- The eye is less sensitive to noise in high resolution bands and in those bands having orientation of 45°
- The eye is less sensitive to noise in those areas of the image where brightness is high or low
- The eye is less sensitive to noise in highly textured areas but, among those, more sensitive near the edges

Embedding

Multiplicative watermarking

- the energy of watermark samples proportional to the corresponding host feature samples

$$f_{w,i} = f_i + \gamma\, w_i\, f_i$$

or

$$f_{w,i} = f_i + \gamma\, w_i\, |f_i|$$

- usually combined with frequency domain watermarking
- it is more difficult to perceive a disturb at a give frequency, if the host asset already contains such a frequency component
- image dependent watermarking

Embedding

- it is more difficult to estimate the watermark by averaging a set of watermarked images
- simultaneously match the invisibility and the robustness constraints
- watermark inserted in the most important parts of the host
- much more difficult to analyze
- hard optimization of detection/decoding
- classical results of digital communication and information theory cannot be used, since they are usually derived under the assumption of additive noise

Detection

- The definition of a reliable procedure to retrieve the information hidden within the host signal is of fundamental importance for the proper development of any data hiding system. This is not an easy task because of the many modifications the host asset may undergo after embedding.

- We derive the detector/decoder structure in some simple cases, dealing with over-simplified channel models, where attacks are either absent or modeled as noise addition. Then we evaluate the error probability of the system for the simplified channel, being aware that a more accurate, experimental, analysis is needed to assess the performance of the systems in more realistic situations.

Detection

Due to the wide variety of attacks and to the difficulties of developing an accurate statistical model of host features, the structure of the detector/decoder is usually derived by considering a simplified channel

- depends on the particular embedding rule
- signal detection in a noisy environment, where noise accounts for both the unknown host signal and the possible presence of attacks

A′ contains the watermark w?

The decision must be taken on the basis of a set of observed

variables coinciding with the set of features f′ extracted from A′.

Detection

- We reformulate the detection problem as a classical hypothesis testing problem.
- We develop an analysis by considering channel systems (e.g., the watermarking of grey level images) where features assume scalar values.
- In order to formulate the detection problem let us assume we want to verify whether an asset A’ contains the watermark or not. The host asset is indicated by A’ instead of A or Aw to make explicit that A’ may coincide neither with the original asset nor with the marked asset.

Detection

- H_0: A′ does not contain w
- H_1: A′ contains w

where H_0 is a composite hypothesis accounting for the following two situations

case 1: A′ is not watermarked

case 2: A′ contains a watermark other than w

Watermark detection amounts to defining a test of the simple hypothesis H_1 versus the composite alternative H_0 that is optimum with respect to a certain criterion.

Detection

Likelihood ratio

In Bayes theory of hypothesis testing the criterion is minimization of risk. Bayes risk is defined as the average of a loss function L_{ij}

- L_{01} is the loss sustained when hypothesis H_0 is in force but H_1 is chosen
- L_{10} is the loss sustained when hypothesis H_1 is in force but H_0 is chosen

Detection

In our case observation variables correspond to the vector f', thus the decision criterion can be a decision rule mapping f' into 0 or 1, corresponding to H_1 and H_0.

R_1 and R_0 are acceptance and rejection regions for hypothesis H_1

$$\varphi(f') = \begin{cases} 1 & f' \in R_1 \quad (H_1 \text{ is in force}) \\ 0 & f' \in R_0 \quad (H_0 \text{ is in force}) \end{cases}$$

Detection

Minimization of Bayes risk leads to a decision criterion based on the likelihood ratio

$$l(f') = \frac{p(f'|H_1)}{p(f'|H_0)}$$

where $p(f'|H_i)$ is the pdf of vector f′ conditioned to hypothesis H_i

Let p_0 and p_1 the a priori probabilities of H_0 and H_1 then the minimum Bayes risk is achieved by letting

$$\varphi(f') = \begin{cases} 1 & \dfrac{p(f'|H_1)}{p(f'|H_0)} > \dfrac{p_0 L_{01}}{p_1 L_{10}} \\ 0 & \text{otherwise} \end{cases}$$

Detection

- The decision rule defined implies that the detector operates by comparing the likelihood ratio against a detection threshold.

- The exact specification of $\varphi(\text{f}')$ requires that the watermark embedding rule is specified and that both the host features and the attack noise are characterized statistically (we will see the example of additive SS watermarking).

Detection

Detection threshold

$$T = p_0 L_{01} / p_1 L_{10}$$

Set T trying to minimize the overall error probability P_e.

P_f probability of revealing the presence of w when w is not actually present (false alarm probability)

$$P_e = p_0 P_f + p_1 (1 - P_d)$$

Detection

To minimize P_e we must set T =1

⇒ $L_{01} = L_{10}$ and $p_0 = p_1$ ⇒ $P_f = 1 - P_d$

The minimum error probability is obtained by letting the probability of missing the watermark and that of falsely revealing its presence equal.

- Bayes decision theory requires that the a priori probabilities of H_0 and H_1 are known (not the case in practical applications)
- with attacks different from white Gaussian noise addition the probability of missing the watermark increases
- in many application false detection probability can not fall below a certain level

Detection

- in most cases it is preferable to adopt a different optimization criterion
- it is preferably to minimize the probability of missing the watermark subject to a constraint on the maximum false detection probability
- given a max allowed false alarm probability try to minimize missed detection probability
  - use likelihood ratio as detection statistic
  - determine threshold according to false alarm probability

Detection

Neyman-Pearson detection criterion

The probability of correctly detecting the watermark is maximized subject to a prescribed limit on P_f

$$\varphi(f') = \begin{cases} 1 & I(f') > T \\ 0 & \text{otherwise} \end{cases}$$

$$P\{ I(f') > T \mid H_0 \} = P_f^*$$

Once T has been fixed, the probability of missing the watermark is

$$1 - P_d = \int_{-\infty}^{T} p(I|H_1)\,dI$$

The performance of a detector based on the Neyman-Pearson criterion are usually expressed by ROC (Receiver Operating Characteristic) curves, in which 1 – P_d is plotted against P_f (or P_d against P_f).

AWGN channel

- In order to go on a model a particular embedding strategy has to be considered
- Additionally, a model to describe host feature and attacks is needed
- We focus on the simplest case: the AWGN channel
  - Additive SS watermarking
  - Gaussian host features (not necessarily true)
  - Additive Gaussian noise as attack

$$f_{w,i} = f_i + \gamma\, w_i + n_i$$

  - Uncorrelated host features (true only for techniques in the frequency domain)
  - f_i and n_i identically distributed random variables (not necessarily true)

AWGN channel

- In the AWGN case optimum detection corresponds to correlation detection

$$\varphi = \frac{1}{n}\sum_{i=1}^{n} f_i' w_i$$

$$\begin{cases} \varphi > T_\varphi \quad H_1 \\ \varphi < T_\varphi \; H_0 \end{cases}$$

- The false and missed detection probability can be computed, and the detection threshold fixed, if the variance and mean of the host feature is known: $\mu_{\rho\,|\,Hi}$ and $\sigma_{\rho\,|\,Hi}$

AWGN channel

$$T_{\varphi} = \sqrt{\frac{2\,\sigma_x^2\,\sigma_w^2}{n}}\,\mathrm{erfc}^{-1}(2P_f^*)$$

Equation that completely characterizes the detector performance

$$(1-P_d)(P_f^*) = \frac{1}{2}\mathrm{erfc}\left(\sqrt{\frac{nSNR}{2}} - \mathrm{erfc}^{-1}(2P_f^*)\right)$$

$$SNR = \frac{\gamma^2\sigma_w^2}{\sigma_x^2}$$

- The detector does not need to know the strength used to embed the watermark
- The embedder can adjust the watermark strength to the asset at hand, trading off between imperceptibility and robustness, without informing the detector of its choice