Lecture slides of the course

Information hiding & secret sharing

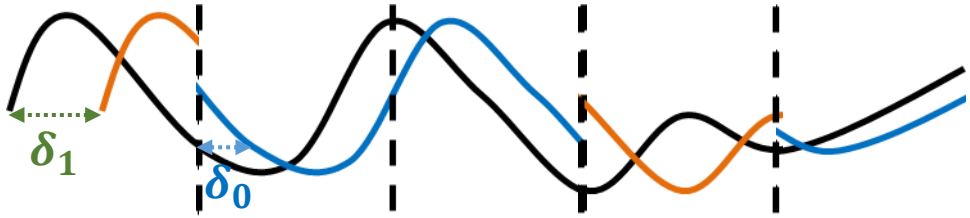
Watermarking

Phạm Trọng Nghĩa

ptnghia@fit.hcmus.edu.vn

The last lecure: hide secret information on audio by echo method

Emdedding

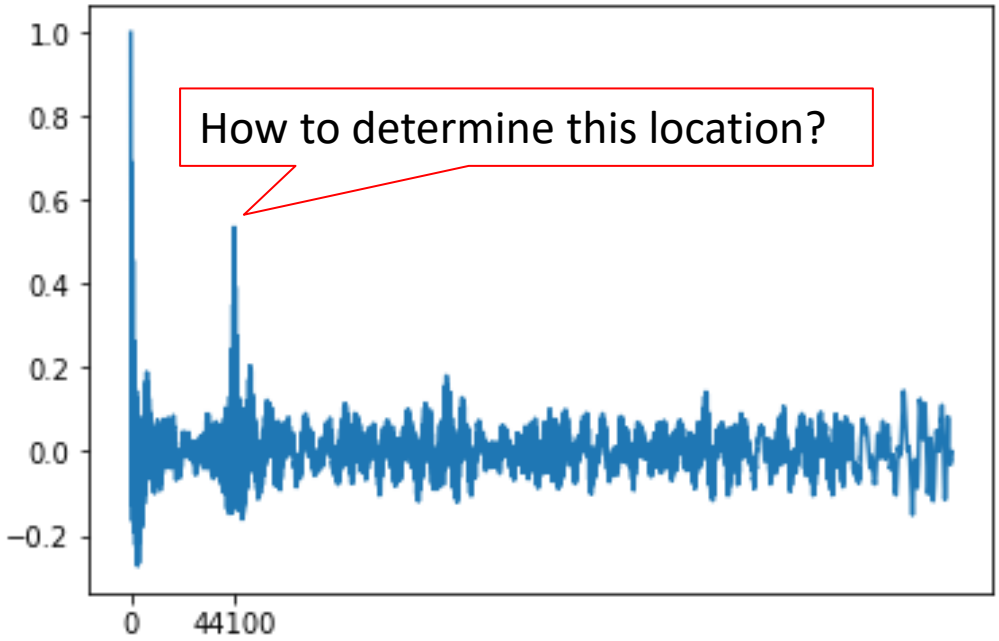


The last lecure: hide secret information on audio by echo method

Extract

Input: stego audio, num segments (number of audio tracks split when embedding)

- Problem: with an audio segment in stego audio, how to identify echo delay\?
- One way is to calculate autocorrelation at different delays...



The last lecure: hide secret information on audio by echo method

Extract

Input: stego audio, num segments (number of audio tracks split when embedding), delta0 (δ_0), delta1 (δ_1)

- Problem: with an audio segment in stego audio, how to identify echo delay at δ_0 or δ_1?
- Given δ_0 and δ_1 is realistic?
  - Possible, because one δ_0 and one δ_1 can be used for many different cover audio
- A simple solution to this problem is to compare the values of autocorrelation at these 2 echo delay δ_0 and δ_1, where have higher autocorrelation is echo position

Course content

- Data/information hiding
  - Steganography ✓
  - Watermarking ← This lecture
- Secret sharing

Problem 1

Let's say you take a good photo and put it on the internet. Others will be able to easily copy your photo, perhaps even say it's theirs, and use it for their own benefit. How to prevent this problem?

One solution

Embed the owner's information in the photo before posting it online.

Sourse image: http://www.aoaophoto.com/watermark-images.htm

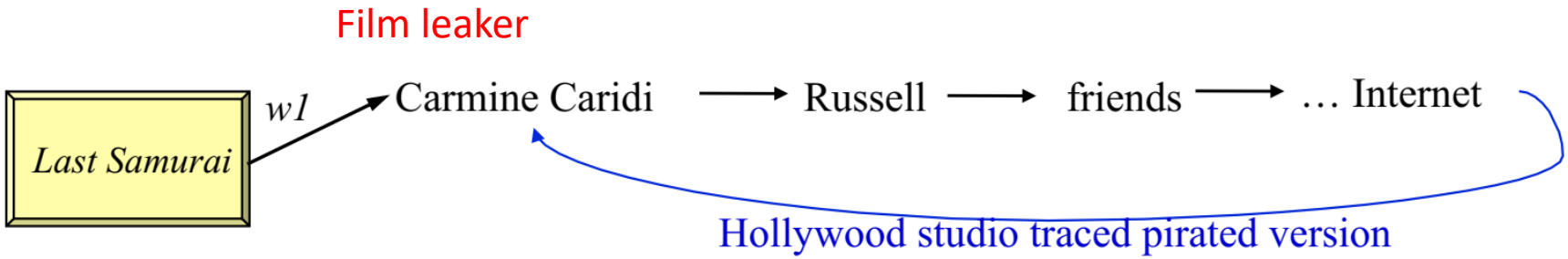
Problem 2

Let's say you are an e-book publisher, you sell multiple copies of a book file to many customers. A customer can arbitrarily create copies, send them to other people; In the end, the result was that the book file appeared on the internet and was freely available to anyone to download. How to prevent this problem?

One solution

When selling a copy to a customer, embed a code specific to that customer. If the book file is leaked online, the publisher can extract the code to know which customer the source of the leak is from.

Real world example



Source:
http://www.cs.ucsb.edu/~htzheng/teach/cs182/schedule/pdf/lecture14.pdf

What is Watermark?

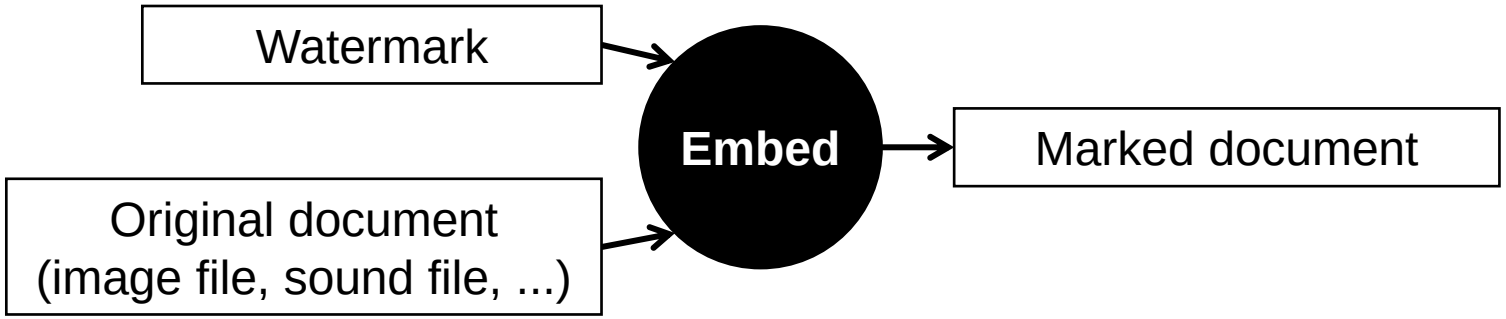
Watermark (digital watermark)) is a "mark" embedded in a data file (text, image, sound, ...) to indicate certain information about the data file (eg, information about the owner) of the data file, or information about the customer who purchased the data file)

How is watermarking different from steganography?

Embedding process watermark

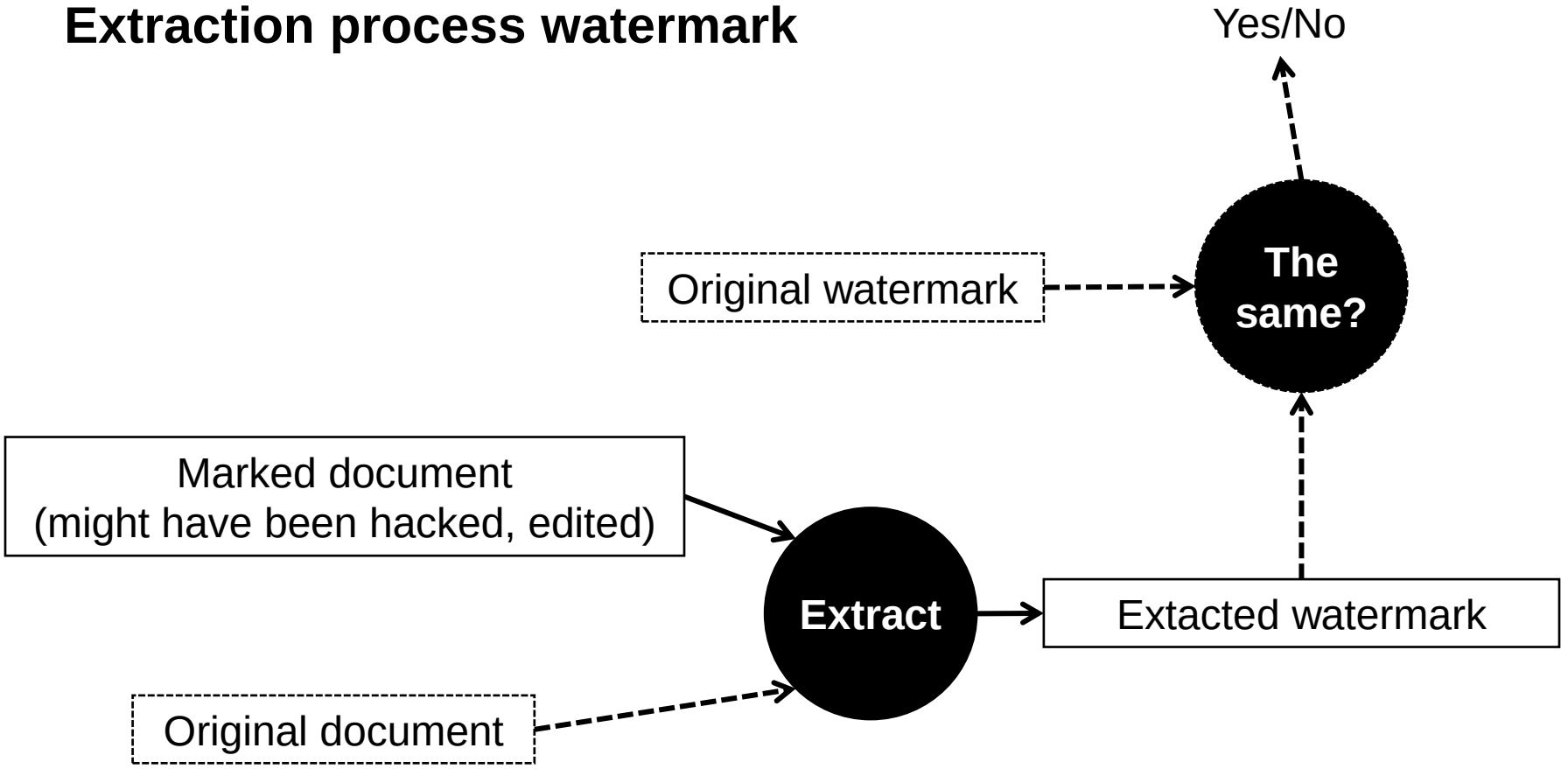
Among the 3 criteria: invisibility – robustness– capacity :

- Steganography emphasize on invisibility and capacity
- Watermarking emphasize on robustness and invisibility



How is watermarking different from steganography?

Extraction process watermark



Dotted parts are optional

Can steganography algorithms be used for watermarking?

Can be used if the steganography algorithm is highly robust (besides high invisibility).

Digital watermarking classification

1. Blind - nonblind
2. Perceptible - imperceptible
3. Private - public
4. Robust - fragile
5. Spatial domain based - frequency domain based

Digital watermarking classification

- A watermarking technique is **blind** if it does not require access to the original unwatermarked data (image, video, audio, etc.) to recover the watermark.
- A watermarking technique is said to be **nonblind** if the original data are needed for the extraction of the watermark
- Nonblind scheme is more robust than the blind
- However, in most applications, the unmodified host signal is not available to the watermark detector.
  - Since the blind scheme does not need the original data, it is more useful than the nonblind one in most applications.

Digital watermarking classification

- A watermark is said to be **perceptible** if the embedded watermark is intended to be visible—for example, a logo inserted into a corner of an image.
- An **imperceptible** watermark is embedded into a host image by sophisticated algorithms and is invisible to the naked eye.
  - It could, however, be extracted by a computer

Digital watermarking classification

- A watermark is **private** if only ***authorized*** users can detect it.
  - Using a private, pseudorandom key to indicates a watermark's location in the host image, allowing insertion and removal of the watermark if the secret location is known.
- Watermarking techniques that allow anyone to read the watermark are called **public**.
  - Public watermarks are embedded in a location known to everyone, so the watermark detection software can easily extract the watermark by scanning the whole image.
- In general, private watermarking techniques are more robust than public ones, in which an attacker can easily remove or destroy the message once the embedded code is known.
- There is also the **asymmetric** form of public watermarking

Digital watermarking classification

- **Robust** watermarks are designed to survive intentional (malicious) and unintentional (nonmalicious) modifications of the image
  - Unintentional modifications include image-processing operations such as scaling, cropping, filtering, and compression.
  - Robust watermarks are usually used for copyright protection to declare rightful ownership.
- **Fragile** watermarks are adopted to detect any unauthorized modification.
  - The slightest modification of the watermarked image will alter or destroy the fragile watermark.

Digital watermarking classification

- In the **spatial domain**, we can simply insert a watermark into a host image by changing the gray levels of some pixels in the host image.
- We can embed the watermark into the coefficients of a transformed image in the **frequency domain**.
  - The transformations include discrete cosine transform, discrete Fourier transform, and discrete wavelet transform.

Digital watermarking classification

- Spatial domain watermarking techniques are usually *less robust to attacks such as compression and added noise*.
- However, they have much *lower computational complexity* and usually can survive a *cropping attack*, which frequency domain watermarking techniques often fail to do.
- Another technique is combining both spatial domain watermarking and frequency domain watermarking for increased robustness and less complexity

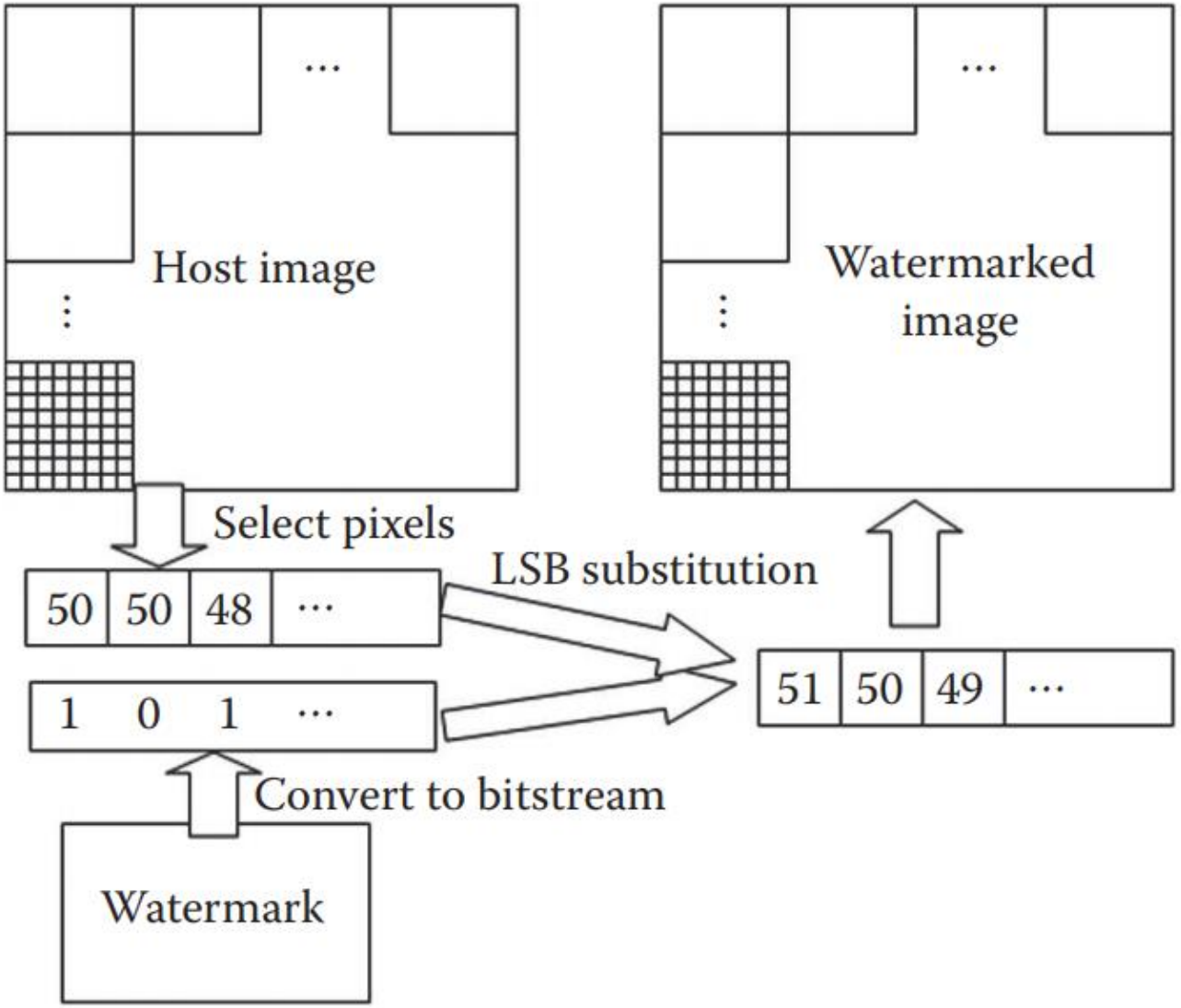
Spatial domain watermarking

- Spatial domain watermarking is the modifying of pixel values directly on the spatial domain of an image
- **Advantage**:
  - In general, spatial domain watermarking schemes are simple and do not need the original image to extract the watermark.
  - They also provide a better compromise between robustness, capacity, and invisibility.
- **Disadvantage**:
  - Not being robust against image-processing operations.
  - The embedded watermark is not distributed around the entire image and the operations can thus easily destroy the watermark.

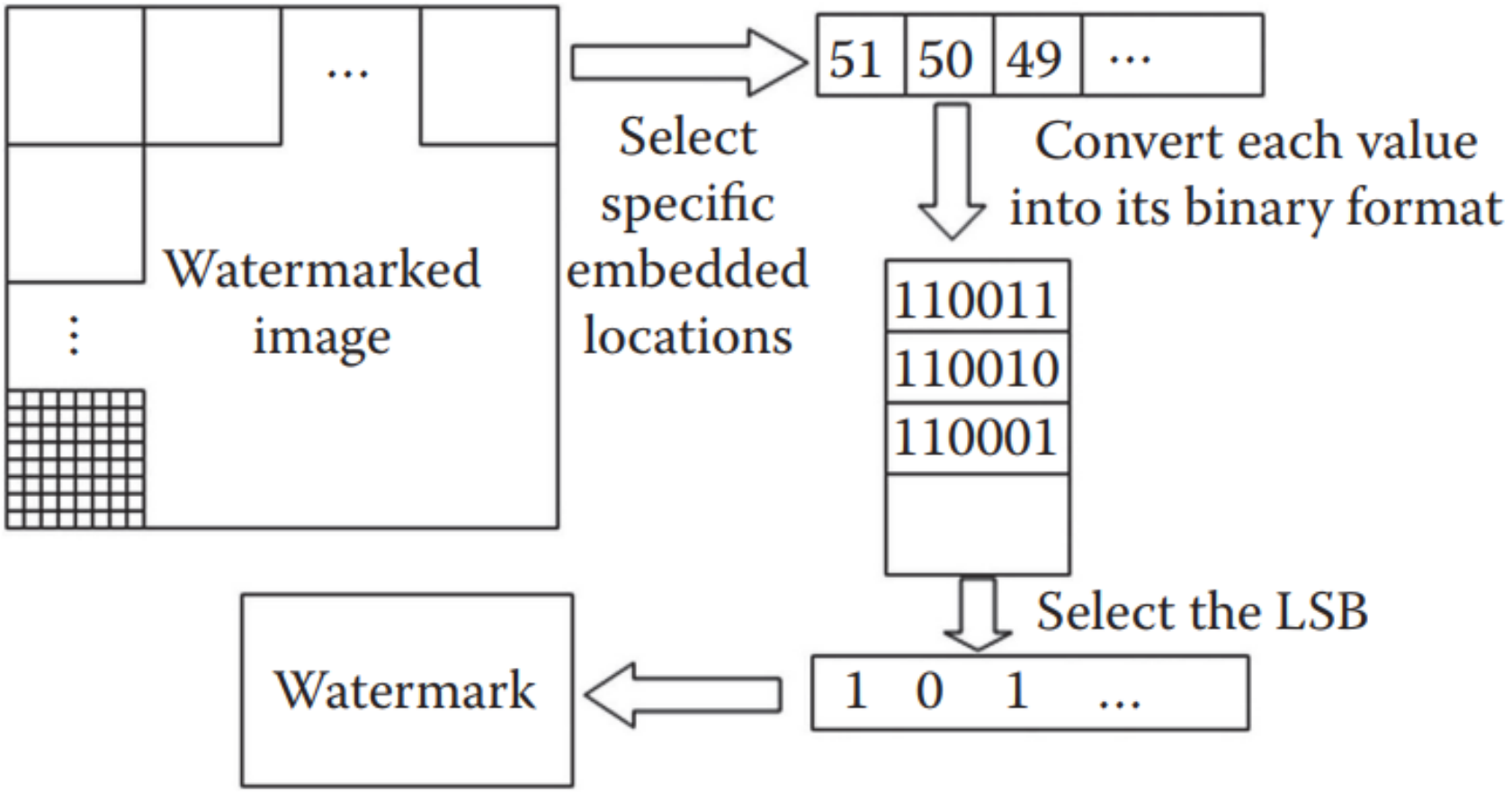
Substitutive Watermarking in the Spatial Domain

- The embedding locations, the specific bits of all pixels, are predefined before watermark embedding.
- Once the recipient obtains the watermarked image, they know the exact locations from which to extract the watermark.
- During the watermark-embedding procedure, the watermark is first converted into a bitstream. Then, each bit of the bitstream is embedded into the specific bit of the selected locations for the host image

Substitutive Spatial domain watermarking – Embed



Substitutive Spatial domain watermarking - Extract



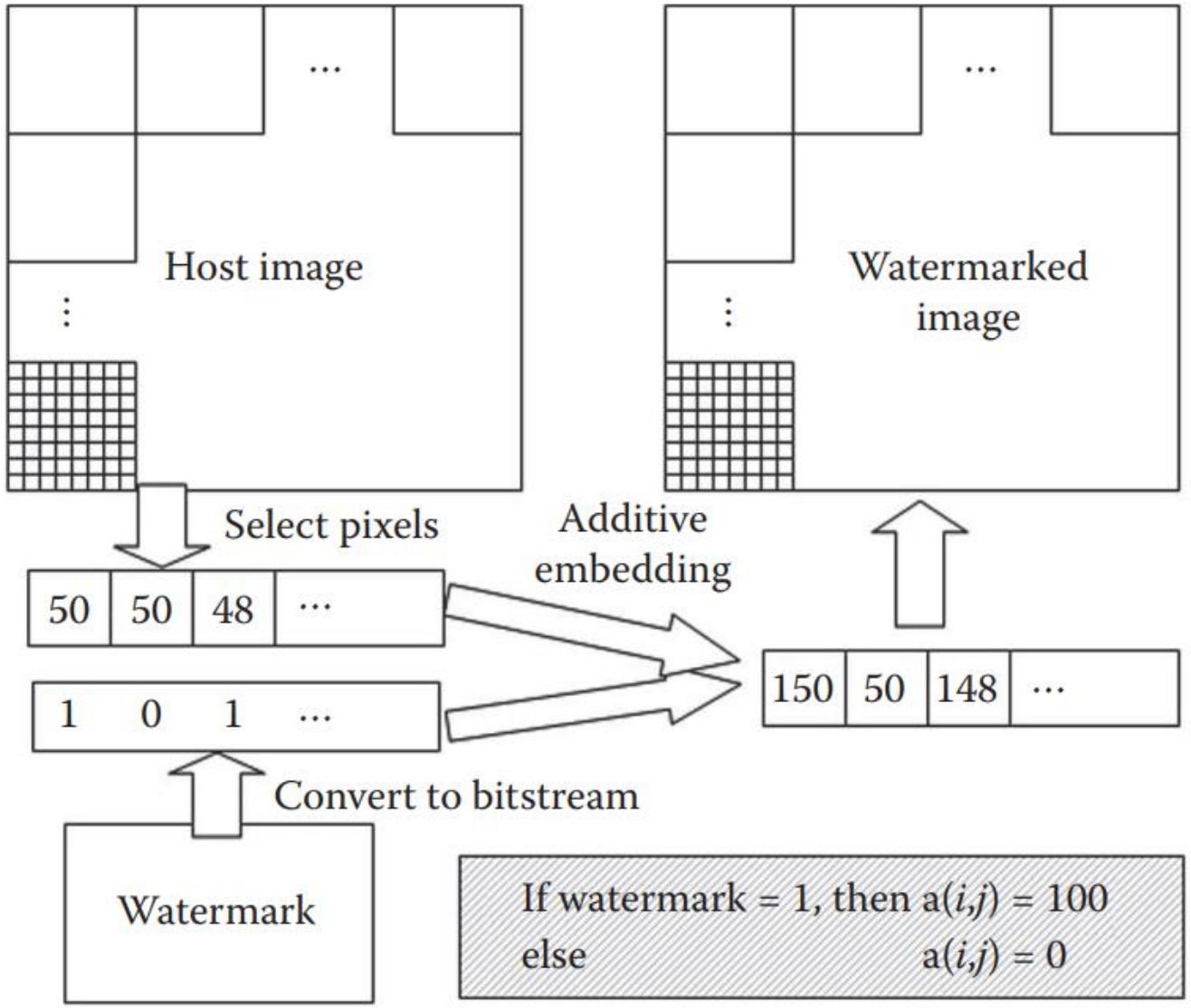
Additive watermarking in the spatial domain

- Adds an amount of watermark value into a pixel to perform the embedding approach.
- If ***H*** is the original grayscale host image and ***W*** is the binary watermark image, then {*h*(*i*,*j*)} and {*w*(*i*,*j*)} denote their respective pixels. We can embed *W* into *H* to become the watermarked image *H** as follows

$$h^*(i,j) = h(i,j) + a(i,j) \cdot w(i,j)$$

- {*a*(*i*,*j*)} is scalling factor

Additive watermarking in the spatial domain



Frequency domain watermarking

- In frequency domain watermarking, we insert a watermark into frequency coefficients of the transformed image via DFT, DCT, or DWT.
- Frequency transforms usually decorrelate the spatial relationship of pixels, the majority of the energy concentrates on the low-frequency components. When we embed the watermark into the low or middle frequencies, these changes will be distributed throughout the entire image.
- When image-processing operations are applied to the watermarked image, they are less affected.
- Therefore, when compared with the spatial domain watermarking method, the frequency domain watermarking technique is relatively more robust

Definition of Good Watermarking

- Invisible
- Robust
  - Common signal processing
  - Common geometric distortions (image and video data)
  - Subterfuge Attacks: Collusion and Forgery
- Universal
- Unambiguous

Definition of Good Watermarking

- **Invisible**
  - The watermark should be perceptually invisible, or its presence should not interfere with the work being protected
- Robust
- Universal
- Unambiguous

Definition of Good Watermarking

- Invisible
- **Robust: watermark must be difficult (hopefully impossible) to remove**
  - Common signal processing
  - Common geometric distortions (image and video data)
  - Subterfuge Attacks: Collusion and Forgery
- Universal
- Unambiguous

Definition of Good Watermarking

- Invisible
- **Robust**
  - **Common signal processing: watermark should still be retrievable even if common signal processing operations are applied to the data. Ex: digital-to-analog and analog-to-digital conversion, resampling, requantization (including dithering and recompression), and common signal enhancements to image contrast and color, or audio bass and treble**
  - Common geometric distortions (image and video data)
  - Subterfuge Attacks: Collusion and Forgery
- Universal
- Unambiguous

Definition of Good Watermarking

- Invisible
- Robust
  - Common signal processing
  - **Common geometric distortions (image and video data): Watermarks in image and video data should also be immune from geometric image operations such as rotation, translation, cropping and scaling**
  - Subterfuge Attacks: Collusion and Forgery
- Universal
- Unambiguous

Definition of Good Watermarking

- Invisible
- Robust
  - Common signal processing
  - Common geometric distortions (image and video data)
  - **Subterfuge Attacks - Collusion and Forgery: the watermark should be robust to collusion by multiple individuals who each possess a watermarked copy of the data. That is, the watermark should be robust to combining copies of the same data set to destroy the watermarks**
- Universal
- Unambiguous

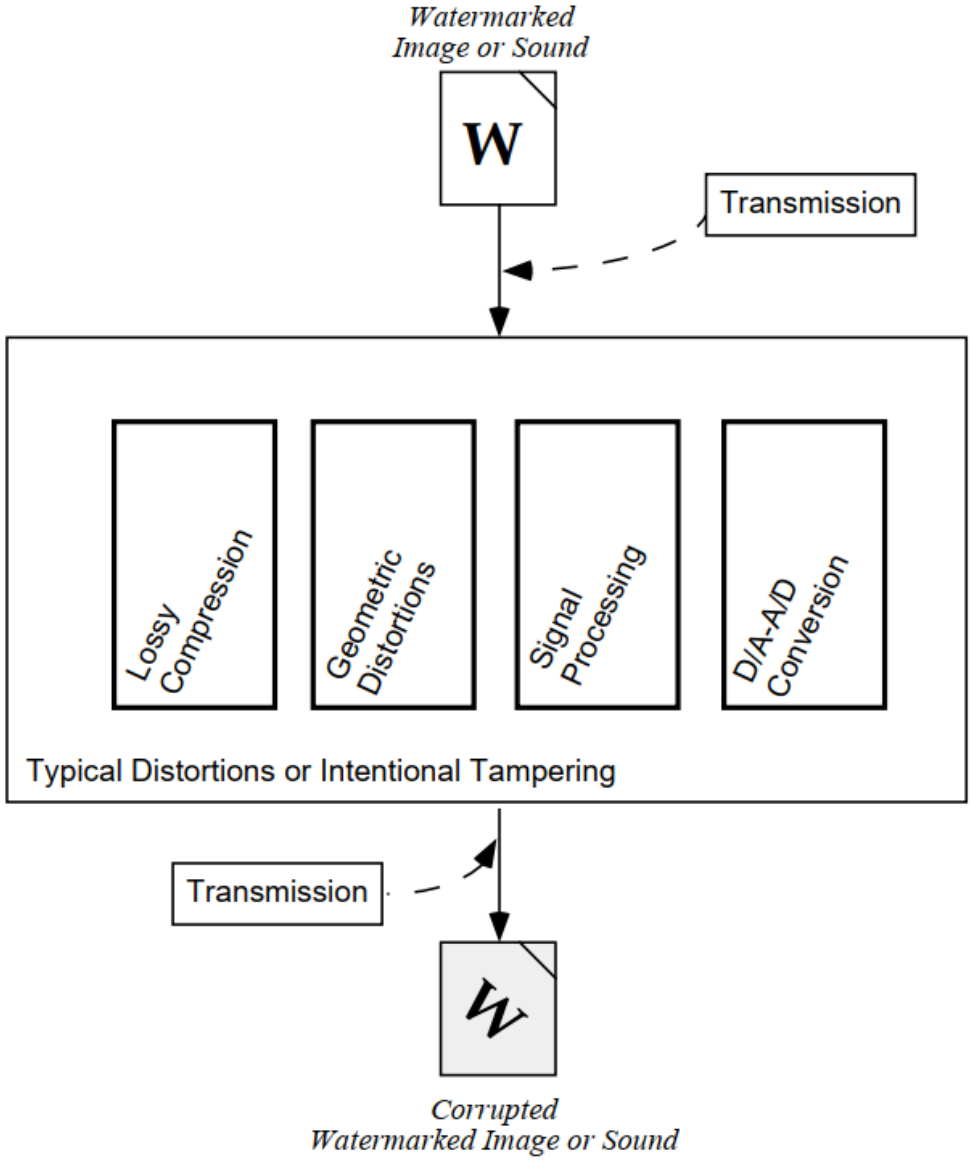
Definition of Good Watermarking

- Invisible
- Robust
  - Common signal processing
  - Common geometric distortions (image and video data)
  - Subterfuge Attacks: Collusion and Forgery
- **Universal:**
  - **The same digital watermark algorithm should apply to all three media under consideration. This is potentially helpful in the watermarking of multimedia products. Also, this feature is conducive to implementation of audio and image/video watermarking algorithms on common hardware**
- Unambiguous

Definition of Good Watermarking

- Invisible
- Robust
  - Common signal processing
  - Common geometric distortions (image and video data)
  - Subterfuge Attacks: Collusion and Forgery
- Universal
- **Unambiguous**
  - **Retrieval of the watermark should unambiguously identify the owner. Further, the accuracy of owner identication should degrade gracefully in the face of attack**

Common signal distortions



Watermarking on photos (or audio) with Cox [1] method

- The watermark should not be placed in **perceptually insignficant regions** of the image or its spectrum since many common signal and geometric processes affect these components.
- For example, a watermark placed in the high frequency spectrum of an image can be easily eliminated with little degradation to the image

Watermarking on photos (or audio) with Cox [1] method

Idea

- Use a watermark as a vector of many elements, each element has a random value and has a small magnitude (consider watermark as an id for a person)
- Embedding a watermark consisting of ***n*** elements into ***n*** important frequency coefficients
- When validating, extract the watermark $w*$ from the image (possibly edited) and calculate the correlation with the original watermark w, if the correlation is greater than the threshold ➔ the image has watermark

[1] Cox et al., Secure Spread Spectrum Watermarking for Multimedia, 1997, Cited by 8248

Watermarking on photos (or audio) with Cox [1] method

Idea

- Use a watermark as a vector of **many** elements, each element has a random value and has a **small magnitude** (consider watermark
- Embedding a watermark consisting of ***n*** elements

into ***n*** **important** frequency coefficients

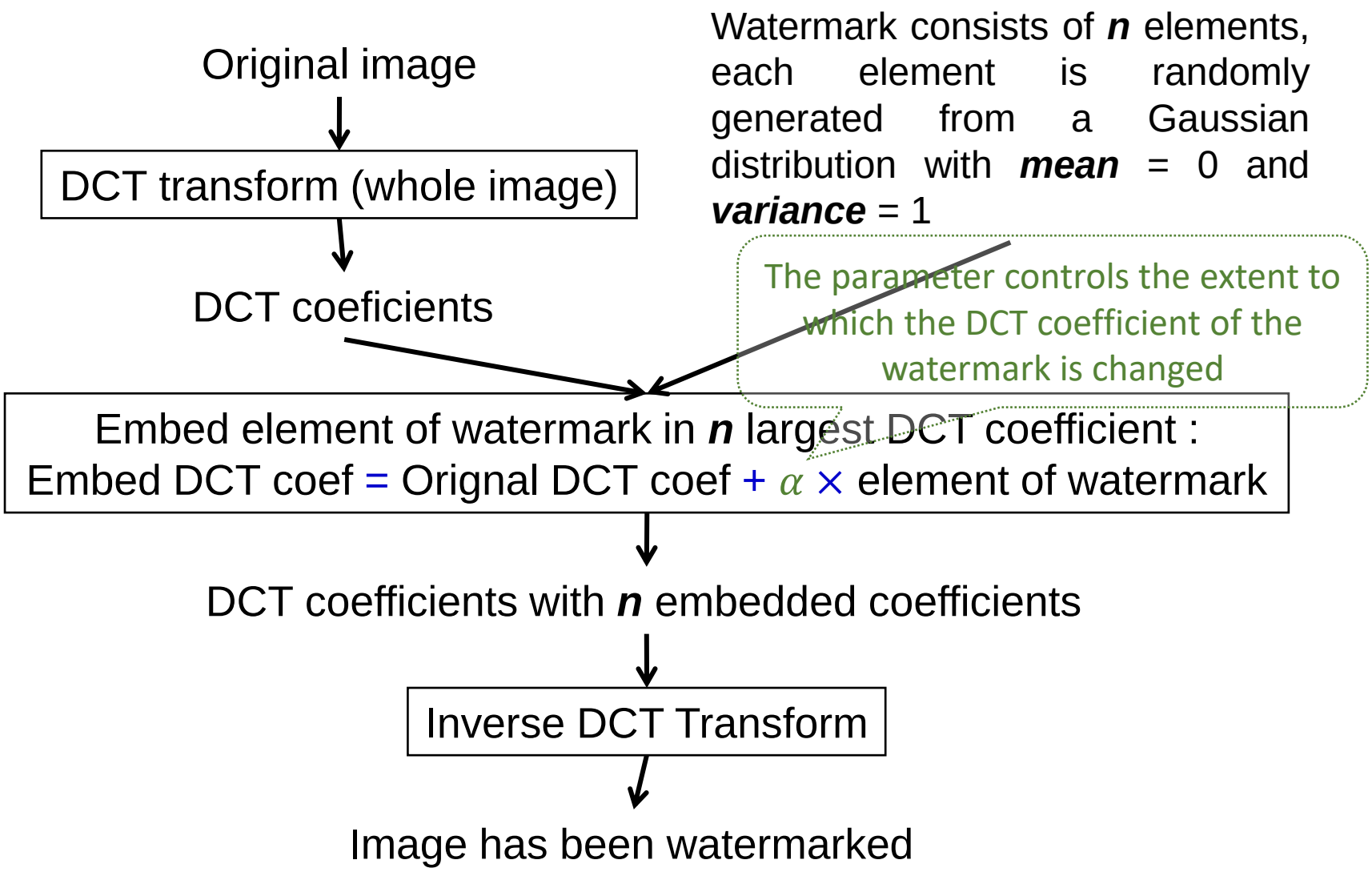
Robustness

Invisible

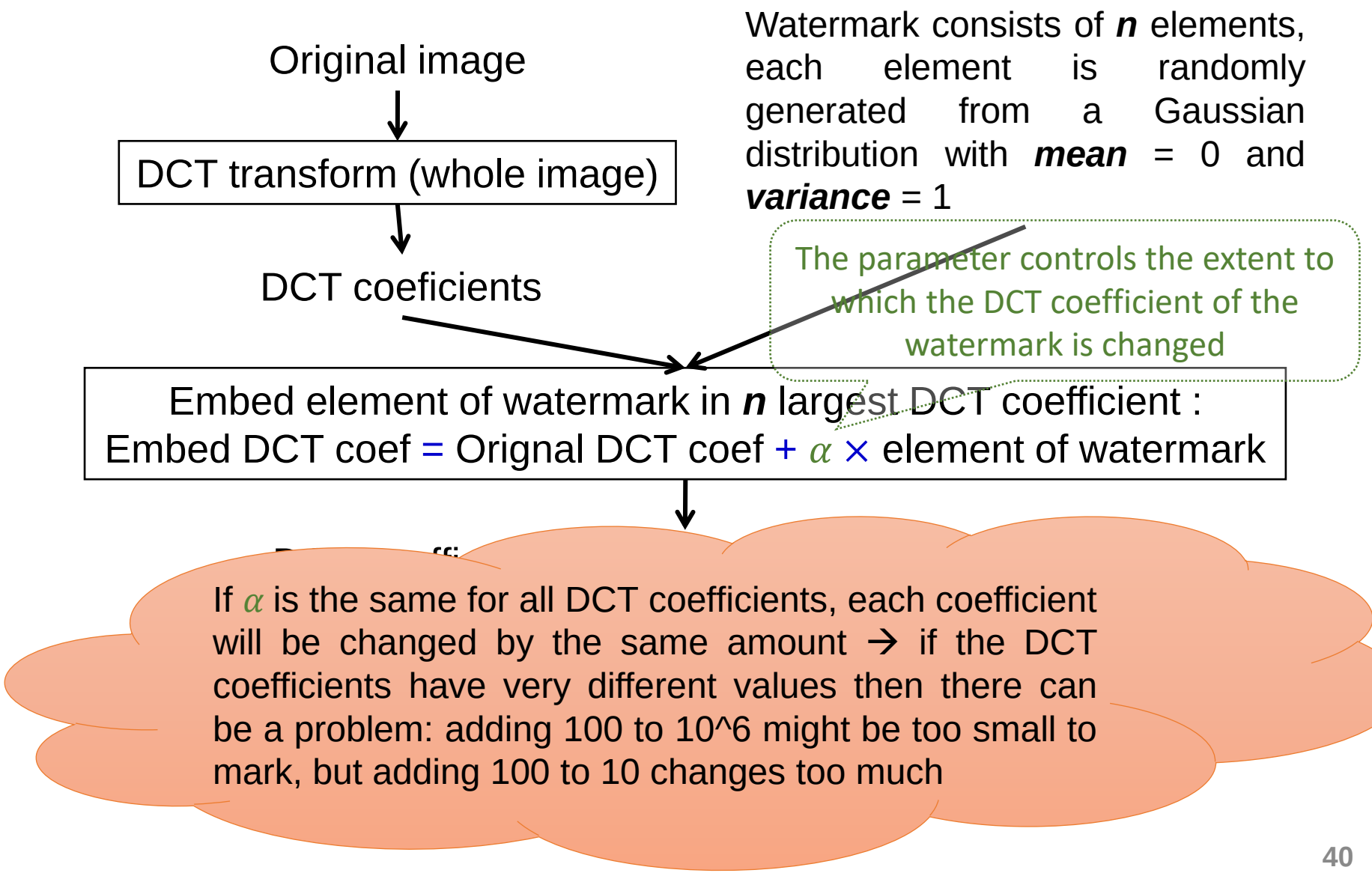
- When validating, extract the watermark $w*$ from the image (possibly edited) and calculate the correlation with the original watermark w, if the correlation is greater than the threshold ➔ the image has watermark

[1] Cox et al., Secure Spread Spectrum Watermarking for Multimedia, 1997, Cited by 8248

Watermarking on images by Cox method: embedding process



Watermarking on images by Cox method: embedding process



Watermarking on images by Cox method: embedding process

Original image

↓

DCT transform (whole image)

↓

DCT coeficients

Watermark consists of ***n*** elements, each element is randomly generated from a Gaussian distribution with ***mean*** = 0 and ***variance*** = 1

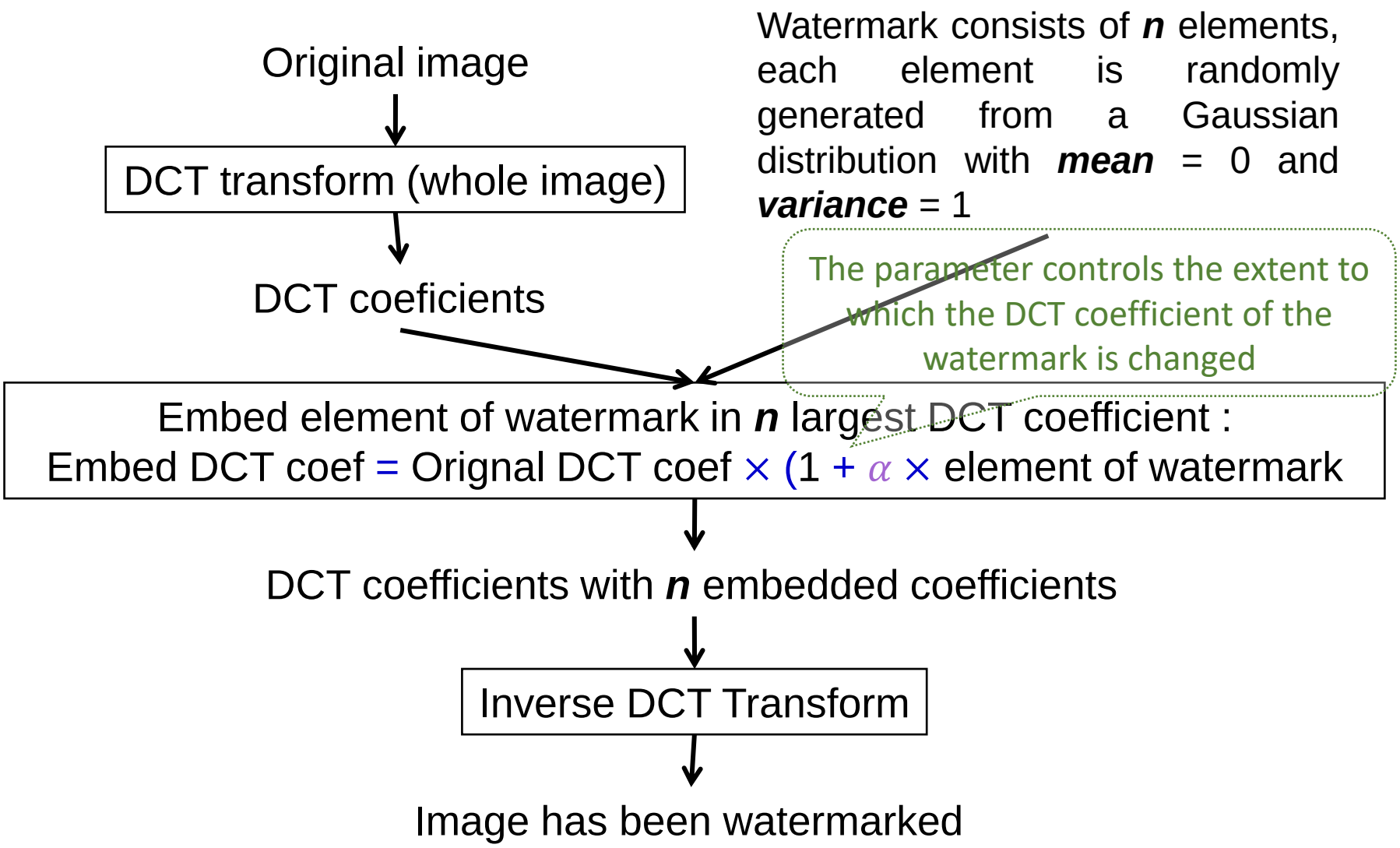
The parameter controls the extent to which the DCT coefficient of the watermark is changed

Embed element of watermark in ***n*** largest DCT coefficient :
Embed DCT coef = Orignal DCT coef + $\alpha \times$ element of watermark

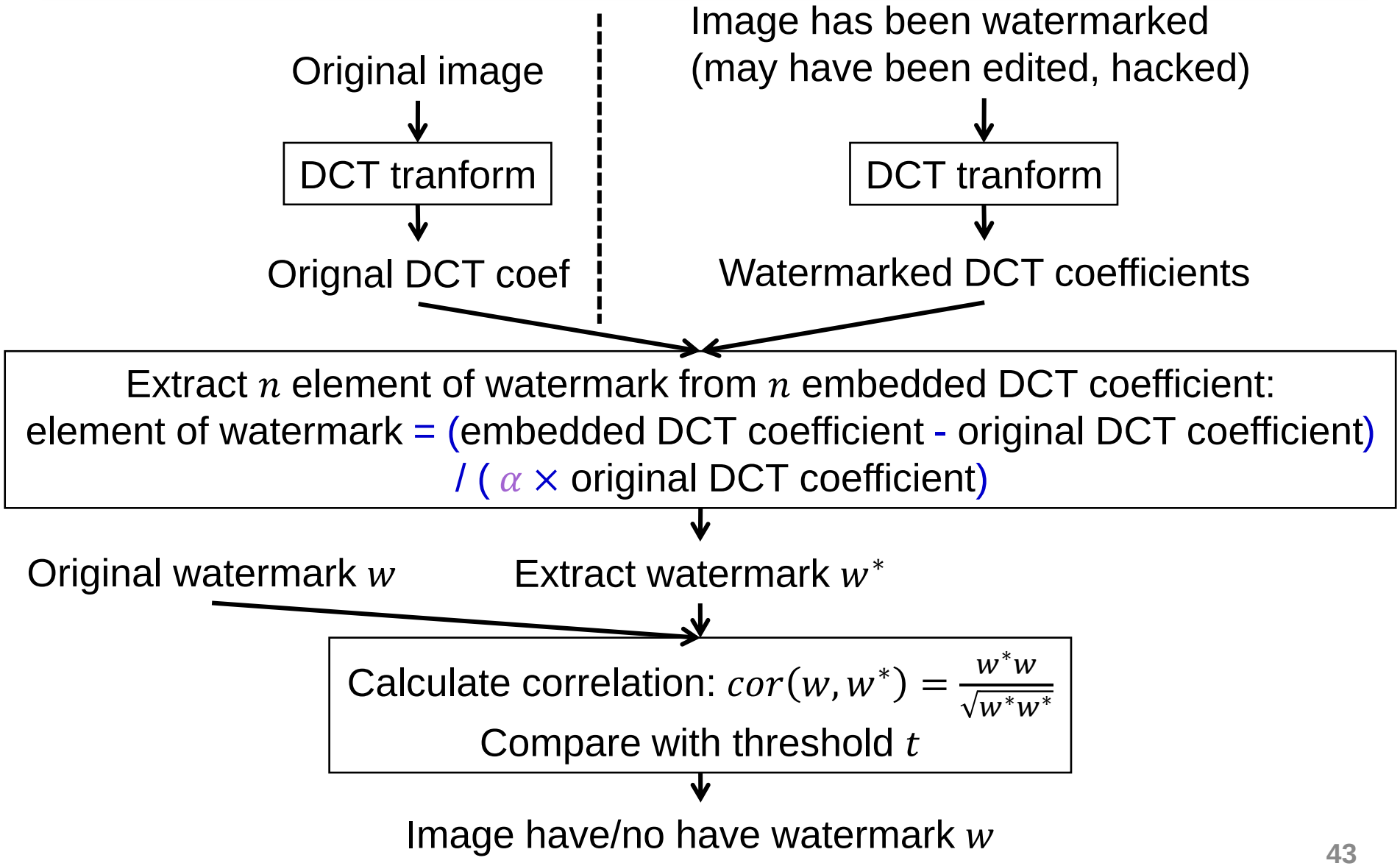
One solution is to use the different α for each DCT coef :

$$\alpha = \alpha \times DCT\ coef$$

Watermarking on images by Cox method: embedding process



Watermarking on images by Cox method: extract process



Watermarking on images using Cox method

How to choose α, watermarking length n, threshold t?

- One way is to conduct experiments to choose the right value
- In paper, author choose:
  - $\alpha = 0.1$
  - $n = 1000$
  - $t = 6$

Demo ...

Reference

- 1. Cox, Ingemar J., et al. "Secure spread spectrum watermarking for multimedia." *IEEE transactions on image processing* 6.12 (1997): 1673-1687.
- 2. Shih, Frank Y. *Digital watermarking and steganography: fundamentals and techniques*. CRC press, 2017.