Lecture 14: Digital Watermarking II

Some slides from Prof. M. Wu, UMCP

Lab2 Demo

Csil

Monday: May 24, 1—4pm
Optional (9:30—11am)

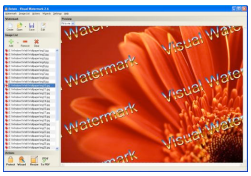
10 minutes per Group

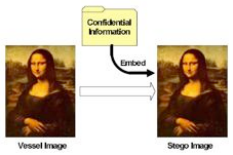
5 Minutes Presentation
5 Minutes Demo

Sign-up Sheet posted outside of my office HFH 1121

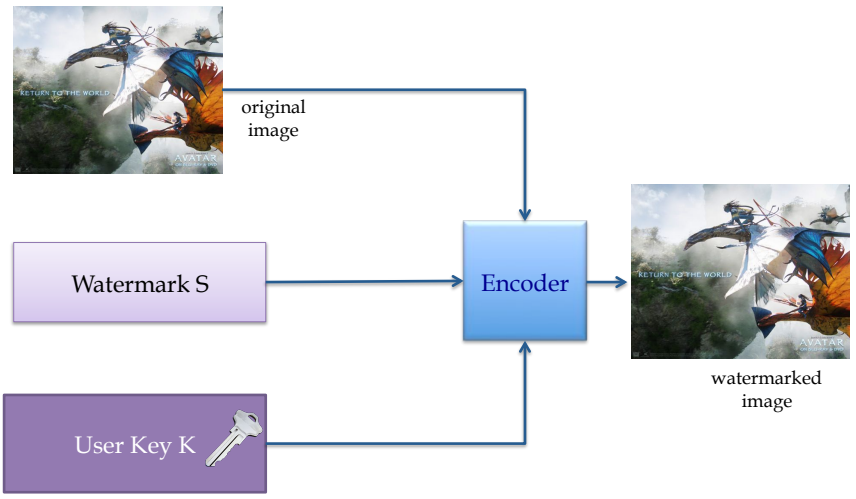
Review: What is a Watermark?

- A watermark is a "secret message" that is embedded into a "cover message"
- Usually, only the knowledge of a secret key allows us to extract the watermark.
- Has a mathematical property that allows us to argue that its presence is the result of deliberate actions.
- Effectiveness of a watermark is a function of its
  - Stealth
  - Resilience
  - Capacity

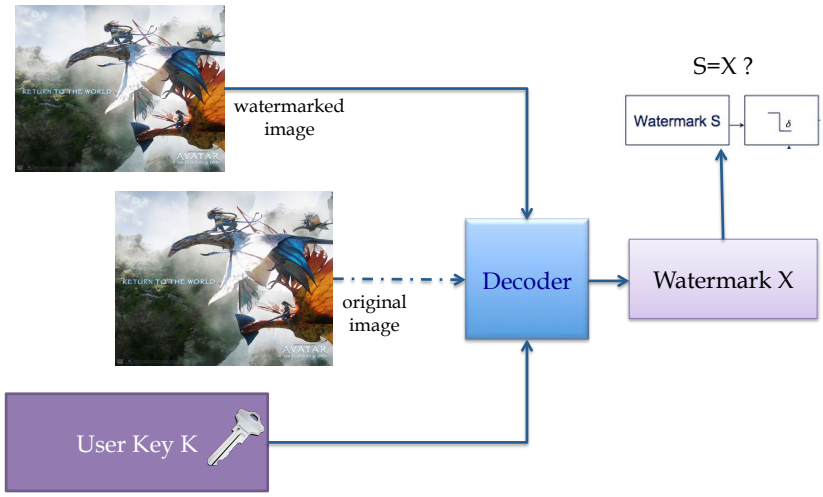




Review: Watermarking Encoding



Review: Watermarking Decoding

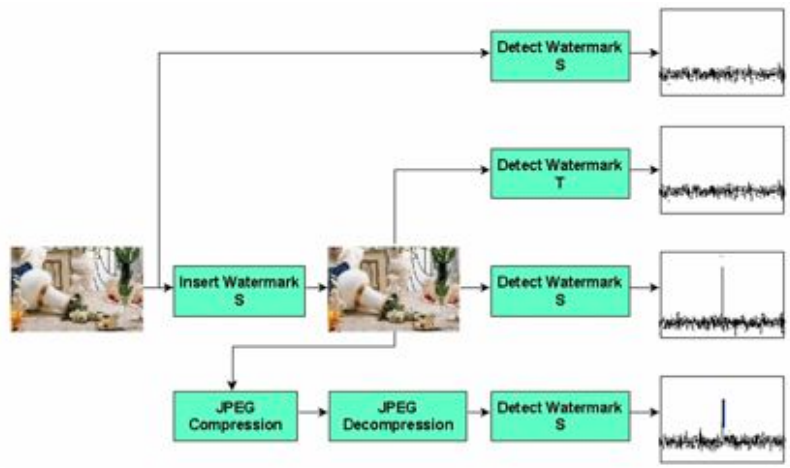


Various Categories of Watermarks

- Based on method of insertion
  - Additive
  - Quantize and replace
- Based on domain of insertion
  - **Transform domain (today)**
  - Spatial domain (last lecture)
- Based on method of detection
  - Private - requires original image
  - Public (or oblivious) - does not require original
- Based on security type
  - Robust - survives image manipulation
  - Fragile - detects manipulation (authentication)

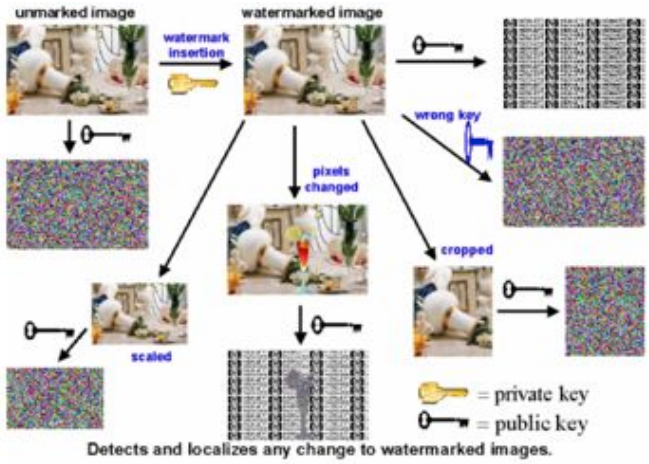
Review: Robust Watermarks

Can still extract the watermark even with editing, noise, or compression etc.



Review: Fragile Watermarks

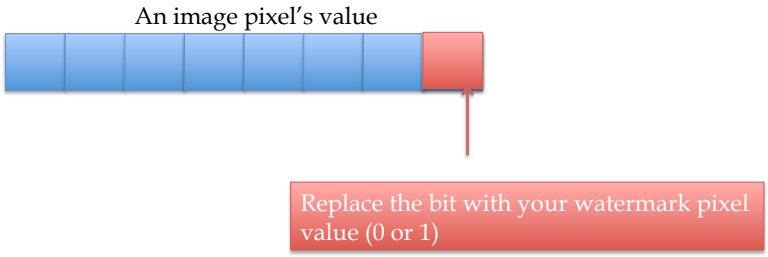
Breaks when the image is altered; used to detect whether the image has been altered.



Detects and localizes any change to watermarked images.

Review: Embedding Fragile Watermarks

- Method 1:
  - Spatial Domain Least Significant Bit (LSB) Modification
  - Simple but not robust



Review: Spatial Domain Robust Watermarking

- Pseudo-randomly (based on secret key) select n pairs of pixels:
  - pair i: a_i, b_i are the values of the pixels in the pair
  - The expected value of sum_i $(a_i-b_i)==0$
- Increase a_i by 1, Decrease b_i by 1
  - The expected value of sum_i (a_i-b_i) now →2n
- To detect watermark, check sum_i (a_i-b_i) on the watermarked image

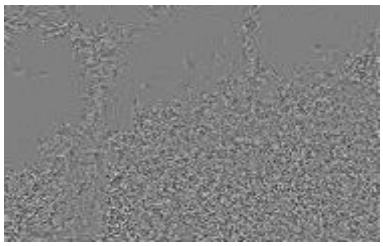
TODAY: FREQUENCY-DOMAIN ROBUST WATERMARKING

Secure Spread Spectrum Watermarking For Multimedia

Cox, Kilian, Leighton, and Shamoon,

IEEE Transactions on Image Processing vol. 6, no. 12, December 1997

An Example



⟺ **10011010 ...**

⇕

© Copyright …

- Embedding domain tailored to media characteristics & application requirement

Spread Spectrum Watermark

- Spread Spectrum == transmits a narrowband signal over a much larger bandwidth
  - the signal energy present in any single frequency is much smaller
- Apply this to watermark:
  - The watermark is spread over many frequency bins so that the (change of) energy in any one bin is very small and almost undetectable
- Watermark extraction == combine these many weak signals into a single but stronger output
  - Because the watermark verification process knows the location and content of the watermark
- To destroy such a watermark would require noise of high amplitude to be added to all frequency bins

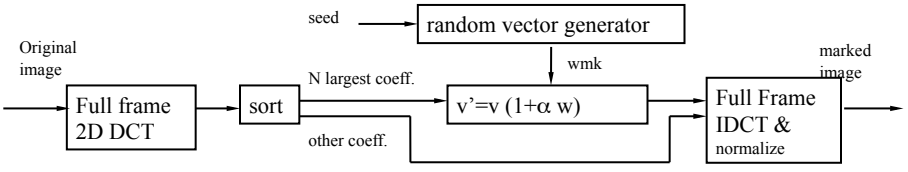
Spread Spectrum Watermark: Cox et al

- What to use as watermark? Where to put it?
  - Place wmk in perceptually significant spectrum (for robustness)
    - Modify by a small amount below Just-noticeable-difference (JND)
  - Use long random noise-like vector as watermark
    - for robustness/security against jamming+removal & imperceptibility
- Embedding $v'_i = v_i + \alpha v_i w_i = v_i (1+\alpha w_i)$

$\alpha = 0.1$

$w_i \sim$ iid, zeromean, unit variance

  - Perform DCT on **entire image** and embed wmk in DCT coeff.
  - Choose **N=1000 largest AC coeff**. and scale $\{v_i\}$ by a random factor



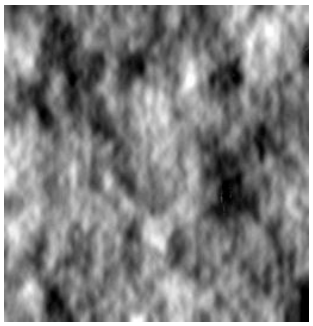
Details: Embedding a Watermark

- Compute the M x M DCT of an M x M gray scale cover image I
- The watermark **W** must be composed of random numbers drawn from a Gaussian distribution **N(0,1)**
  - $N(\mu, \sigma^2)$ denotes a normal distribution with mean μ, and variance σ^2
- Embed a sequence of watermark: $W = w_1, w_2, \ldots, w_n$, according to N (0,1), into the **n largest** magnitude DCT coefficients X_i, excluding the DC component)
  - Type I: $X_i' = X_i + \alpha w_i$, i=1....n
  - Type II: $X_i' = X_i(1 + \alpha w_i)$ i=1….n
- Now compute the inverse DCT to obtain the watermarked image I'
  - In general α=0.1, n=1000

Watermarking Example by Cox et al.

Original

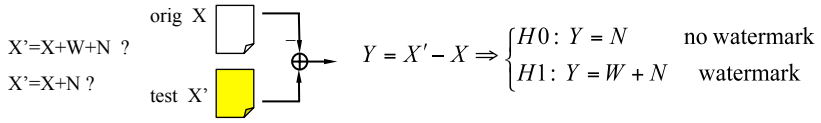
Cox
whole image DCT
Embed in 1000 largest coeff.



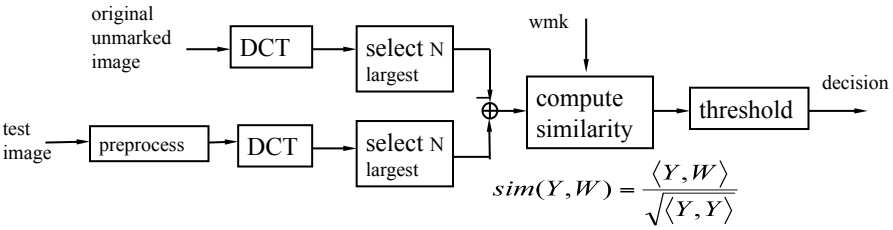
Difference between marked & original

Cox et al's Scheme (cont'd): Detection

- Subtract original image from the test one before feeding to detector ("non-blind detection")



- Correlation-based detection
  - a correlator normalized by |Y| in Cox et al. paper



Details: Detection Steps

- Compute the DCT of the watermarked (and possibly attacked) cover image I*
  - Need original image and compute its own DCT values
  - Find the n largest AC coefficients from the original image
- Extract the watermark W
  - For Add-SS: $y_i = (x_i^* - x_i)/\alpha$
  - For Mult-SS: $y_i = (x_i^* - x_i)/\alpha x_i$
- Evaluate the similarity of Y and W using sim

$$sim(Y,W) = \frac{\langle Y,W \rangle}{\sqrt{\langle Y,Y \rangle}}$$

- If sim (Y, W) > T, a given threshold, the watermark W exists

Performance of Cox et al's Scheme

- Robustness

Distortion	none	scale 25%	JPG 10%	JPG 5%	dither	crop 25%	print-xerox-scan
similarity	32.0	13.4	22.8	13.9	10.5	14.6	7.0

threshold = 6.0 (determined by setting false alarm probability)

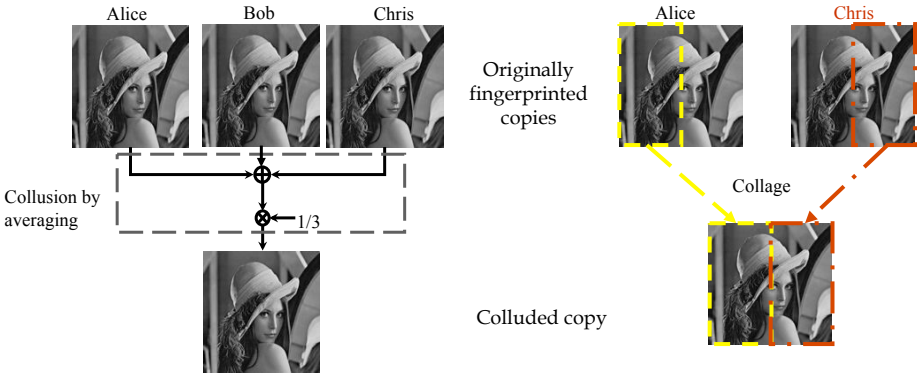
  - (claimed) scaling, JPEG, dithering, cropping, "printing-xeroxing-scanning", multiple watermarking
  - No big surprise with high robustness
    - equivalent to sending just 1-bit {0,1} with $O(10^3)$ samples

Summary: Spread Spectrum Embedding

- Main ideas
  - Place wmk in perceptually significant spectrum (for robustness)
    - Modify by a small amount below Just-noticeable-difference (JND)
  - Use long random vector of low power as watermark to avoid artifacts
    (for imperceptibility, robustness, and security)
- Cox's approach
  - Perform DCT on entire image & embed wmk in large DCT AC coeff.
  - Embedding: $x'_i = x_i + \alpha x_i w_i = x_i (1+\alpha w_i)$
  - Detection: subtract original and perform correlation w/ wmk

Collusion Attacks by Multiple Users

- Collusion: A cost-effective attack against multimedia fingerprints
  - Users with same content but different fingerprints come together to produce a new copy with diminished or attenuated fingerprints
  - Fairness: *Each colluder contributes equal share through averaging, interleaving, and nonlinear combining*



Your Lab3

Implement the spread spectrum watermark embedding in matlab

[60pt] (**estimated time: 2--3 hours**)

- Read original image (download lena.jpg here), make it 8-bit grayscale)
  - **rgb2gray(), uint8()**
- Generate watermark vector w of length n (e.g., n = 1000)
  - **randn()**
- Apply 2D DCT transform on the **entire image, not each macroblock (dct2())**
- Take the n largest AC coefficients x
- Generate watermarked coefficients x' by x' = x * (1 + a * w)
  w is the corresponding watermark component, a = 0.1
- Apply 2D IDCT on the new DCT coefficients (original DC, new x' and the rest AC coefficients) **(idct2())**
- Compare the original and watermarked image
  - Compute the PSNR
- Repeat the above with different n (100, 200, 500, 1000, 1500)
- Plot a figure of PSNR vs. n

Your Lab3 cont

Implement the spread spectrum watermark detection in matlab

[40pt] Detect/extract watermark (**estimated time, 2--3 hours**)

- apply 2D DCT on the image to be tested
- extract the n largest coefficients (Hint: use the original image to identify the location of these n coefficients)
- Subtract the corresponding n DCT value of the original image,
  - $y_i = (x_i^* - x_i)/\alpha x_i$
- Compute the similarity, and use a threshold of **6** to check with a particular watermark W is present

$$sim(Y,W) = \frac{\langle Y,W \rangle}{\sqrt{\langle Y,Y \rangle}} \qquad \langle Y,W \rangle = \sum_i y_i \cdot w_i$$

Due Tue. June 1, 11:59PM via turnin lab3, Individual project, no grouping

Your Lab3 cont

Examine Robustness in matlab

Bonus [25pt] Check robustness (**estimated time: 2-3 hours**)

- Add noise, alter your image (JPEG compression with quality 25 and 50), and see if you can detect the original watermark
  - Both Noise and JPEG compression steps can be found in the watermarking_demo.zip from the course website

```
outfilename='temp.jpg';
imwrite(wimg,outfilename,'Quality',25);
wcimg=imread(outfilename);
```

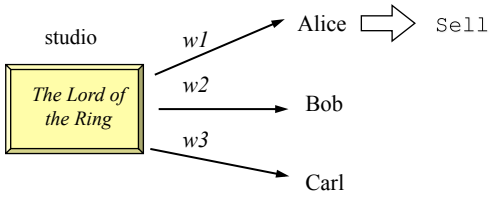
```
noisyimg = imnoise(wimg,'gaussian');
```

- Produce collusion attacks
  - Generate three different watermarked (w1,w2,w3) images following the same manner, take the average of the three to create a new image, and detect whether any watermark (w1,w2, w3) is present
- Try the above with different n values and report your findings

FINGERPRINTING

Robust Wmk Application for Tracing Traitors

- Leak of information as well as alteration and repackaging poses serious threats to government operations and commercial markets
  - e.g., pirated content or classified document

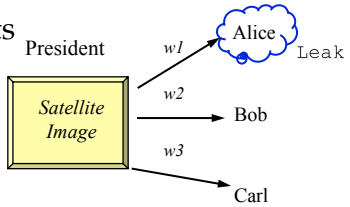


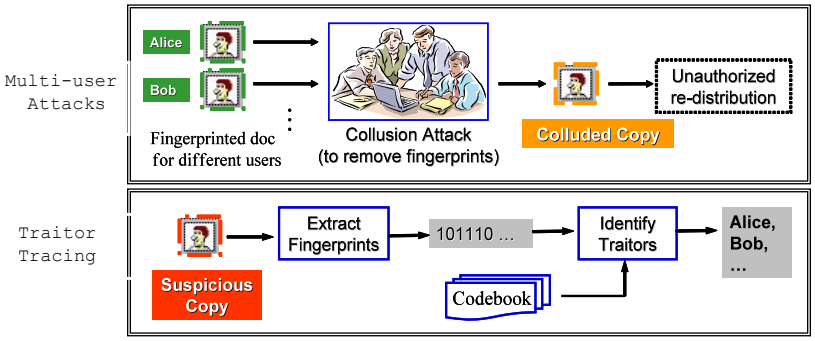
- Promising countermeasure: robustly embed digital fingerprints
  - Insert ID or "fingerprint" (often through conventional watermarking) to identify each user
  - Purpose: deter information leakage; digital rights management (DRM)
  - Challenge: imperceptibility, robustness, tracing capability

Embedded Fingerprint for Tracing Traitors

Insert special signals to identify recipients

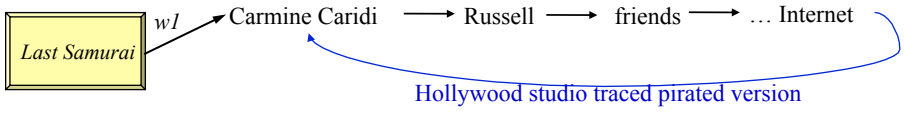
- Deter leak of proprietary documents
- Complementary protection to encryption
- Consider imperceptibility, robustness, traceability
- Attacks mounted by single and multiple users





Case Study: Tracing Movie Screening Copies

- Potential civilian use for digital rights management (DRM)
  - *Copyright industry – $500+ Billion business ~ 5% U.S. GDP*
- Alleged Movie Pirate Arrested (23 January 2004)
  - A real case of a successful deployment of 'traitor-tracing' mechanism in the digital realm
  - Use invisible fingerprints to protect screener copies of pre-release movies



http://www.msnbc.msn.com/id/4037016/

Summary of Digital Watermarking

- Widely used to protect digital media
- Many different forms (still an active research area today)
  - Spatial vs. Frequency
  - Robust vs. Fragile

Schedule for the Next 2 weeks

- Mon (May 24): No class, Demo Session at Csil
- Wed. (May 26): Discussion on lab3
- Mon (May 31): No class, holiday ☺
- Wed. (June 2): Final review