Image Data Compression

Introduction to watermarking and steganography
Examples and common terminology

Example: watermarks (WMs) embedded in paper money
- Hidden from view (or non-obstructing) in normal use
- Recovered via a special process (holding up to light)
- Contains information related to object (bill authenticity)
- The presence of a watermark may be known to user

Example: a secret message written with milk on top of a non-secret (cover, or decoy) letter written with ink
- Hidden message is unrelated to the “decoy” letter
- The presence of the hidden message itself is secret (otherwise called “overt embedded communication”)
- From Greek “steganos” = “covered”

Watermarking of electronic signals:
Work: a specific signal, such as an image, audio or video record (e.g., song in MP3)
Content: set of all Works (e.g., all audio music) that may be WM’ed
Medium: means of representing, transmitting or recording the content (e.g., a CD)
Cover Work: original signal before modification (called message embedding)

Watermarking: imperceptibly altering a Work to embed a message about that Work
Steganography: undetectably altering a Work to embed a secret message
Steganalysis: detection whether secret steganographic communication is taking place

Book on subject: [Cox et al, ‘08]
# Categorization of information hiding

<table>
<thead>
<tr>
<th>Existence of message is hidden</th>
<th>Cover Work-dependent message</th>
<th>Cover Work-independent message</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covert watermarking</td>
<td><strong>Example:</strong> In 1981, M. Thatcher marked copies of secret documents with unique word spacing patterns to find a cabinet minister who had leaked information.</td>
<td>Steganography <strong>Example:</strong> SALT-II nuclear treaty described sensors on missile silos, reporting if a silo is occupied. Both USSR and USA investigated if sensors could transmit other information.</td>
</tr>
<tr>
<td>Overt watermarking</td>
<td><strong>Example:</strong> Web-sites of some museums provide high-quality digital images of the collection, with a warning that each image is watermarked to protect it against piracy or reproduction.</td>
<td>Overt embedded communication <strong>Example:</strong> In late 1940s, a time code at 800Hz frequency was embedded in radio broadcast. The code was inaudible, and only communicated the current time to various automatic devices.</td>
</tr>
</tbody>
</table>

![Diagram showing the process of information hiding](Diagram)
Possible uses of digital watermarking and steganography

**Watermarking:**
- Internet and high-capacity digital recording devices facilitate unauthorized copying
- Cryptography provides the protection in transit, but not after delivery
- Watermarking is a complement to cryptographic protection, never removed during use, may be designed to survive content transformations (re-encoding, format changes, etc.)

**Steganography:**
- Electronic communications are susceptible to eavesdropping and interventions
- Security and privacy can be addressed by cryptographic tools, or by anonymous remailers
- However, encryption is not hidden, and the presence of the communication is obvious
- Steganography can be used also in cases when the message encryption is prohibited
- Cryptanalysis aims to establish whether communication is taking place (e.g. to prevent criminal activities or to identify members of an organization)
Some specific applications of digital watermarking

- Broadcast monitoring
  air time verification, re-broadcasting control, ads control
  active vs passive monitoring: index complexity reduction
- Owner identification
  legal copyright notice, owner contact information
- Proof of ownership
  key in central repository, digital “negative”, asymmetric
  identification of original / derived work
- Transaction tracking (fingerprinting of copies)
  e.g. identification of leaking party at Oscar Award previews
- Content authentication
  tampering detection, localized [semi]fragile digital signature
- Copy / Record / Playback control
  e.g. DVD copy protection
- Device control
  copy prevention marks, ads indicator, traffic info on FM radio
- Legacy enhancement
  digital signals over analog networks, lyrics in MP3

Desired WM properties and characteristics:
- Imperceptibility: WM must not ruin the aesthetics of Cover Work
- Inseparability: WM cannot be removed by converting, re-formatting, etc.

Example of (relatively poor) watermarking as owner identification: the complete Lena image and its (usually omitted) copyright notice
Quantitative metrics of watermarking systems

- **Embedding effectiveness**
  probability that output is identified as watermarked immediately after embedding (may be <100%) compromise between effectiveness and fidelity; determined analytically or with large DB of Works

- **Fidelity**
  how imperceptible WM is in Work; perceptual similarity between original and watermarked Works (possibly after additional degradation of both due to delivery); based on some perceptive model

- **Robustness**
  how well WM survives common signal processing operations: spatial / temporal filtering, lossy compression, printing / scanning, geometric distortions etc.

- **Data payload**
  number of bits a watermark encodes per unit of time or per Work; N-bit WM encodes $2^N+1$ possible detector outputs (one bit always encodes very presence of WM)

- **Blind [public] or informed [private] detection**
  whether original Work is needed for successful detection

- **False positive rate**
  probability to erroneously detect a missing WM per detector run (fixed Work, random WMs)

- **Security**
  ability to resist hostile attacks, such as unauthorized removal (elimination, masking, collusion) / embedding (forgery) = active attacks, unauthorized detection = passive attack

- **Use of secret watermark key (similar to cipher key)**

- **Cost**
  deployment of embedders / detectors, computational load, real-time requirements etc.
Recall the primary goal of steganography:
conceal the fact that the covert communication is present within innocuous communication

Properties of a WM system irrelevant for steganography:
• Embedding effectiveness: N/A, due to freedom to choose a suitable Cover Work
• Fidelity: N/A, since the steganalyzing party normally has no access to the original Work
• Blind / informed extraction: N/A, one usually assumes that the original Work is not available
• Robustness: N/A, noise etc. not a big issue for modern digital communication

Properties important for steganography:
• Embedding capacity: maximum theoretically possible number of embedded bits
• Steganographic capacity: max payload hidden without artifacts, so that the detection is improbable
• Embedding efficiency: number of embedded bits per unit of distortion
• Robustness against system / blind / targeted steganalysis
  Based on method weakness (implementation fault, insufficient stego keyspace), statistical properties common to all methods, detectability of messages embedded with a specific method.
• Statistical undetectability
  Is it hard to notice the presence of a message? Usually (loosely!) quantified in terms of statistical anomalies, based on some statistical model of relevant Works.
• False alarm rate: tradeoffs characterized with Receiver Operating Characteristic (ROC) curves
• Security: resistance to passive (observation), active (obstruction), malicious (impersonation) attacks
• Use of stego keys
  Algorithm is assumed known, embedding is controlled by a secret key. Schemes can be symmetric or asymmetric; as a rule, key length does not influence the security as much as the length of crypto keys
Small digression: communication systems

Reminder: generic image compression + communication system

- Source
  - Transform
  - Lossless coding

- Sink
  - Inverse transform
  - Decoding

Codec

Inject channel-adjusted redundant information (error protection)

Exploit, remove injected redundancy

Simple communication system:

- Input message: \( m \)
- Channel encoder
- \( \vec{x} = \{x_1, x_2, \ldots, x_N\} \), \( \sum x_i^2 \leq p \)
- Simplest channel: additive white Gaussian noise, \( \vec{y} = \vec{x} + \vec{n}, \vec{n} \sim N(\vec{0}, \sigma^2) \)
- Channel
- Noise
- Output message: \( m_n \)
- Channel decoder
- Goal: maximize likelihood that the detected message is identical to the original
Secure communication systems

**Communication with encryption:**

<table>
<thead>
<tr>
<th>Input message</th>
<th>Encryptor</th>
<th>Channel encoder</th>
<th>Channel decoder</th>
<th>Decryptor</th>
<th>Output message</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m$</td>
<td>Encryption key</td>
<td>$\vec{x}$, $\vec{y}$</td>
<td>$\vec{n}$</td>
<td>Decryption key</td>
<td>$m_n$</td>
</tr>
</tbody>
</table>

**Communication with key-based channel coding:**

<table>
<thead>
<tr>
<th>Input message</th>
<th>Channel encoder</th>
<th>Channel decoder</th>
<th>Output message</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m$</td>
<td>Encoding key</td>
<td>Decoding key</td>
<td>$m_n$</td>
</tr>
</tbody>
</table>

Goal: secrecy of messages, messaging layer. Example: RSA encryption

Goal: guaranteed delivery of signals, transport layer. Example: spread-spectrum radio
Communication-based models of watermarking

Watermarking with a simple informed detector:

- **Watermark encoder**
  - Input message $m$
  - Watermark key
  - Original cover work $c_o$
  - Watermark $w_a$
  - Watermark embedder
  - Output message $m_n$
  - Watermark detector
  - Noise $n$
  - Original cover work $c_w$
  - Watermark decoder

- **Process Steps**
  1. Message mapped to *added pattern* $w_a$ (could be via intermediate *message pattern* $w_m$)
  2. $w_a$ is added to cover work $w_o$ to produce *watermarked work* $c_w$ (i.e. blind embedder: ignores properties of cover work)
  3. Further processing adds noise $n$ (could be more complex: compression, attacks etc.)
  4. If original work is subtracted, then the communication process is identical to simple communication model with additive noise
Watermarking with a blind detector:

- In blind detector, cover work is just another kind of noise
- One goal: maximize similarity between the input and output messages
- Another possible goal: learn how exactly watermarked work was processed
Embed 1-bit message $m$: 

$$\tilde{c} = \tilde{c}_o + \alpha (2m-1) \cdot \tilde{w}_r$$

Detect message via **linear correlation** (scalar product):

$$z_{lc}(\tilde{c}, \tilde{w}_r) = \frac{1}{N} \tilde{c} \cdot \tilde{w}_r = \frac{1}{N} \sum_{x,y} c[x,y] \cdot w_r[x,y]$$

Effect from noise:

$$\tilde{c} = \tilde{c}_o \pm \alpha \cdot \tilde{w}_r + \tilde{n},$$

$$z_{lc}(\tilde{c}, \tilde{w}_r) = z_{lc}(\tilde{c}_o, \tilde{w}_r) + z_{lc}(\tilde{n}, \tilde{w}_r) \pm \alpha \cdot z_{lc}(\tilde{w}_r, \tilde{w}_r)$$

If reference pattern has zero mean, unit variance, then:

$$m_n = \begin{cases} 
1, & z_{lc}(\tilde{c}, \tilde{w}_r) > \tau \\
no, & |z_{lc}(\tilde{c}, \tilde{w}_r)| < \tau \\
0, & z_{lc}(\tilde{c}, \tilde{w}_r) < -\tau 
\end{cases}$$
Testing simple WM system with DB of images

- DB of 4000 images
- Strength $\alpha = 1$
- Threshold $\tau = 0.7$ results in false positive probability of $\sim 10^{-4}$
- Performance highly dependent on reference image!
- E.g. low-pass filtering results in much poorer separation, higher FP probability

Working and useful WM system; however, only optimal if the cover work and noise are drawn from Gaussian distribution, susceptible to certain attacks
**Watermarking as communication with side information**

**Idea:** allow embedder to examine cover work before generating watermark

- Can even subtract cover work completely!

- [Shannon, ‘58]: communication with side information at the transmitter
- Can modify embedder to guarantee 100% embedding efficiency (losing in fidelity)
Fixed correlation embedding

\[
\tilde{c} = \tilde{c}_o + \alpha (2m - 1) \cdot \tilde{w}_r, \quad \alpha = \frac{\tau_{goal} - z_{lc}(\tilde{c}_o, \tilde{w}_r)}{z_{lc}(\tilde{w}_r, \tilde{w}_r)}
\]

Percentage of images

Detection value (linear correlation)
Geometric models of watermarking

**Media space:** multi-dimensional space of all works

Fidelity distance function extremely difficult to formalize; depends on human perception. Simplest case - MSE:

\[
D_{\text{MSE}}(\tilde{c}_1, \tilde{c}_2) = \frac{1}{N} \| \tilde{c}_1 - \tilde{c}_2 \|^2
\]

\[
= \frac{1}{N} \sum_{x,y} (c_1[x,y] - c_2[x,y])^2
\]

Some perceptual distance functions are asymmetric, result in units of JND – just noticeable difference

True distribution of works in media space is usually unknown, effect of transmission / noise / attacks also unknown 😊 (but modeled anyway)
Effects of blind and fixed-correlation embedding

Alternative common detection region definitions:

- **Normalized correlation**: angle in N-dim space
  \[
  z_{nc}(\vec{c},\vec{w}_r) = \frac{\vec{c} \cdot \vec{w}_r}{|\vec{c}| \cdot |\vec{w}_r|} = \cos(\vec{c},\vec{w}_r)
  \]

- **Correlation coefficient**: subtract mean first
  (= angle between N-1-dimensional projections)
  \[
  z_{cc}(\vec{c},\vec{w}_r) = z_{nc}(\vec{c} - \langle \vec{c} \rangle,\vec{w}_r - \langle \vec{w}_r \rangle)
  \]

Plane \( \langle \vec{x} \rangle = \sum x[i] = 0 \)

Plane projections on plane
Marking space (cf. transform-based compression)

- **Media space** (e.g. pixel values) not always convenient for watermarking
- **Extractor** transforms work to more convenient representation (e.g. frequencies, wavelets)
- Unlike image compression, not looking for more compact representation!

---

**Generic watermark embedder**

1. **Original cover work** → **Vector in media space** → **Watermark extractor** → **Vector in marking space** → **Simple WM embedded** → **Inverse extractor** → **Vector in media space** → **Watermarked work** → **...**

**Generic watermark detector**

1. **Input work** → **Vector in media space** → **Watermark extractor** → **Vector in marking space** → **Simple WM detector** → **Detected message** → **...**