Introduction to Visible Watermarking

IPR Course: TA Lecture
2002/12/18
NTU CSIE R105
Outline

- Introduction
- State-of-the-Art
- Characteristics of Visible Watermarking Schemes
- Attacking Visible Watermarking Schemes
- Discussions and Conclusions
Classifying Watermarking Schemes

Data hiding

Steganography

- Imperceptible data embedding
  - Non-robust data embedding
  - Robust data embedding

- Visible data embedding

Watermarking

- Imperceptible watermarking
- Robust watermarking
  - Visible watermarking

- Fragile watermarking
Visible Watermarking

• IPR protection schemes for images and video that have to be released for certain purposes
• Unobtrusive copyright patterns can be recognized on embedded contents
# Invisible Watermarking v.s. Visible Watermarking

<table>
<thead>
<tr>
<th></th>
<th>Invisible Watermarking</th>
<th>Visible Watermarking</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Watermark Perceptibility</strong></td>
<td>Imperceptible distortion</td>
<td>Visibly Meaningful pattern</td>
</tr>
<tr>
<td><strong>Robustness</strong></td>
<td>Intentional attacks and common signal processing</td>
<td>User-intervention based watermark removal</td>
</tr>
<tr>
<td><strong>Protection</strong></td>
<td>Passive</td>
<td>Active</td>
</tr>
<tr>
<td><strong>Extraction</strong></td>
<td>Explicit extraction module</td>
<td>Direct viewing</td>
</tr>
<tr>
<td><strong>Current Research Status</strong></td>
<td>Hot</td>
<td>Only few papers</td>
</tr>
</tbody>
</table>
Requirements of Visible Watermarking

- **Perceptibility of host image details**
  - Contents should not be rendered useless after being visibly watermarked

- **Visibility of watermark patterns in embedded contents**
  - No explicit watermark extraction techniques are required

- **Robustness**
  - Difficult to remove unless exhaustive and expensive human interventions are involved
A General Model of Visible Watermarking

\[ I' = K_1 * I + K_2 * W \]
\[ D(E_I(I'), E_I(I)) < \text{Threshold}_I \]
\[ D(E_W(I'), E_W(W)) < \text{Threshold}_W \]

- \( I' \): the watermarked content
- \( I \): the un-watermarked original content
- \( W \): the watermark pattern
- \( K_i \): the weighting factor
- \( D \): a distance function measuring the perceptual difference of its two parameters
- \( E_i \): image feature extraction operators
- \( \text{Threshold}_I \): the largest allowable distortion of image details that observers can tolerate and, at the same time, the signature of \( I \) can be maintained.
- \( \text{Threshold}_W \): the largest allowable distortion of the embedded watermark pattern that the copyright information can be clearly recognized.
State-of-the-Art

An approximately uniform color space is used, such as the CIE 1976 (L\*u\*v\*) space and the CIE 1976(L\*a\*b\*)-space, so amounts of brightness increasing and decreasing are perceptually equal for a fixed change occurred everywhere in the color space.

**Definitions**

- $Y_{n,m}$ and $Y'_{n,m}$: the brightness values of each pixel in the unmarked original and the watermarked image.
- $Y_w$: the brightness of the “scene white”.

The Scheme Proposed by G. Braudaway et al

\[
Y'_{n,m} = Y_{n,m} + \frac{(\mu_{n,m} - \mu_r)}{|\mu_A - \mu_r|} \frac{Y_w}{38.667} \left(\frac{Y_{n,m}}{Y_w}\right)^{2/3} \Delta L^*
\]
Other Enhancing Schemes

- [Meng and Chang]
  - The same embedding model is extended to the DCT domain by simple statistic model approximation for the convenience of processing directly in the MPEG-compressed domain.

- [Kankanhalli et al]
  - Local features related to the degree of distortion tolerances, such as edge locations, texture distributions and luminance sensitivity, are taken into consideration so that more unobtrusive watermarked images can be generated.
  - Simple statistics of block-DCT coefficients are calculated and analyzed to decide the watermark embedding energy of each block.
    - Edge integrity will be preserved, in these approaches, since the edge information is essential to maintain the image quality.
    - And the energy of the embedded watermark is larger in highly textured areas than in smooth ones due to different noise sensitivity.
    - In additions, the watermark energy of mid-gray regions is also smaller than other areas since the noises are more visible against a mid-gray background.

- [S. P. Mohanty et al]
  - in addition to the visibly embedded watermark, a fragile invisible watermark is also adopted to check if the visible watermark is altered or not.
Important observations (1/4)

- Attacking visible watermarking scheme means successfully recover the watermarked area.

- Implication:
  - Similar image processing techniques can be adopted
    - Image recovery
    - Object removal
Important observations (2/4)

- To clearly recognize the copyright patterns, the contours of embedded patterns must be preserved.

- Implication:
  - An attacking scheme is effective if
    1. The pattern is completely removed
    2. The shape is seriously distorted without seriously degrading visual quality.
Important observations (3/4)

- The perceptibility of the host image details within watermarked area depends on the preservation of edge information.

Implication:

- Available information while attacking
  - Surrounding pixels around watermarked area.
  - Edge information within watermarked area is available while attacking.
Important observations (4/4)

- The robustness lies in the inevitability of exhaustive and expensive labors.

- Implication:
  - Only minimum user intervention should be adopted during attacking
    - User selection of watermarked areas
Averaging Attacks

- Refill the watermarked areas by averaging surrounding pixels.
  - Good approximations for small areas.
  - Blurring effects across object boundaries.
Image Inpainting


\[ I_t^n(i, j) = \left( \delta L^n(i, j) \cdot \frac{N(i, j, n)}{|N(i, j, n)|} \right) \nabla I^n(i, j) \]

- Image inpainting
  - is an **iterative** image recovery technique.
  - prolongs the approaching isophotes into damaged areas.
  - successfully reconstructs the edges of damaged areas.
Basic Inpainting Attacks

- Attacks against visible watermarking are regarded as common image recovery problems.
- Good results can be obtained for areas composed of thin copyright patterns, but areas composed of thick patterns cannot be successfully recovered.
General Attacks

- Classifying flat areas within watermarked area by analyzing remaining edge information of host images
- Directly extend colors of surrounding flat areas into watermarked areas
Further Improvement
Experimental Results (I)
Experimental Results (II)
Experimental Results (III)