Digital Watermarking and Steganography

by Ingemar Cox, Matthew Miller, Jeffrey Bloom, Jessica Fridrich, Ton Kalker

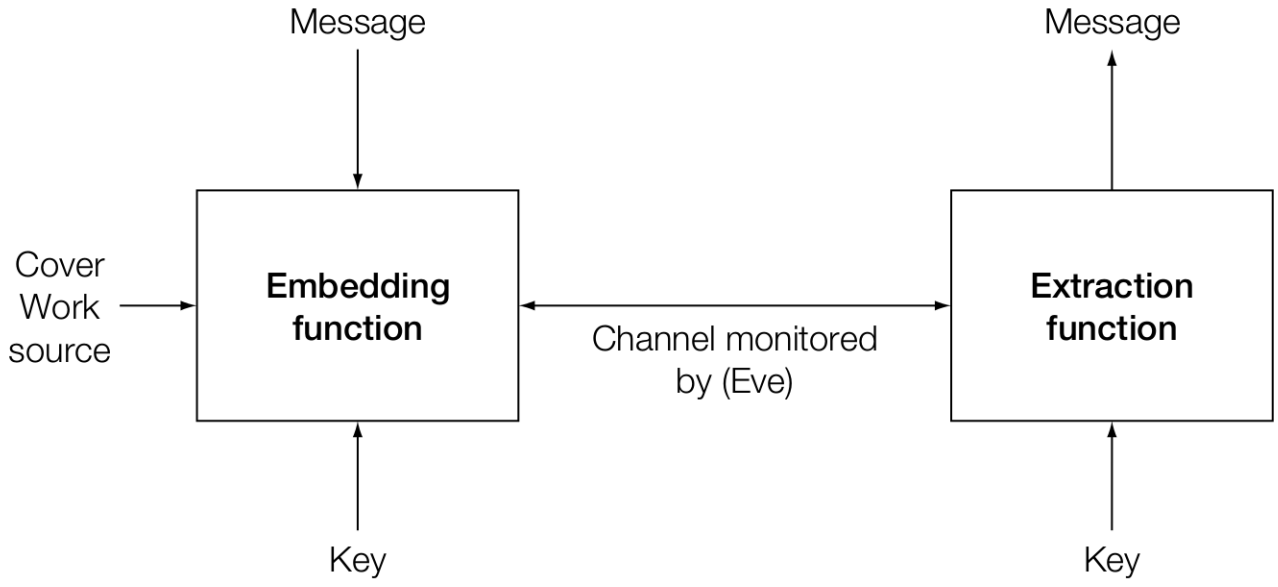
Chapter 12. Steganography

Lecturer: Jin HUANG

Difference to Watermark

- Imperceptible: watermark.
- Undetectable: steganography.

The Model


Message
Message
Cover
Work
source
Embedding
function
Channel monitored
by (Eve)
Extraction
function
Key
Key

The Warden

The warden is part of the channel.

- **Passive**
- Active
- Malicious: trying to impersonate Alice or Bob or otherwise tricking them.

Embedding

The cover work is

- Preexisting, and will not be modified: cover lookup.
- Generated, and will not be modified: cover synthesis.
- Preexisting and modified: cover modification.

Look up

payload 依.

- Labeling work by messages.
- Deliver the messages by sequence of transmission.

Example

- 1024 songs for 10-bit message.
- 1024 sequential transmissions lead to $10k$-bit.

Synthesis

Creates the stego Work without recourse to a cover Work.

British spies in Wold War II

- Source: a big book of conversations.
- By selecting different phrases from the book.

Packed but nature sequence of look up.

Modification

- Type and magnitude of change.
- Location of change
  - Sequential
  - (Pseudo) random: pseudo-random walk.
  - Adaptive: informed. (背景不改，又改学生多的地方)

The Secret Key

Shared between Alice and Bob

- Seed the pseudo-random walk.
- Seed the noise signal.

The First Attempt

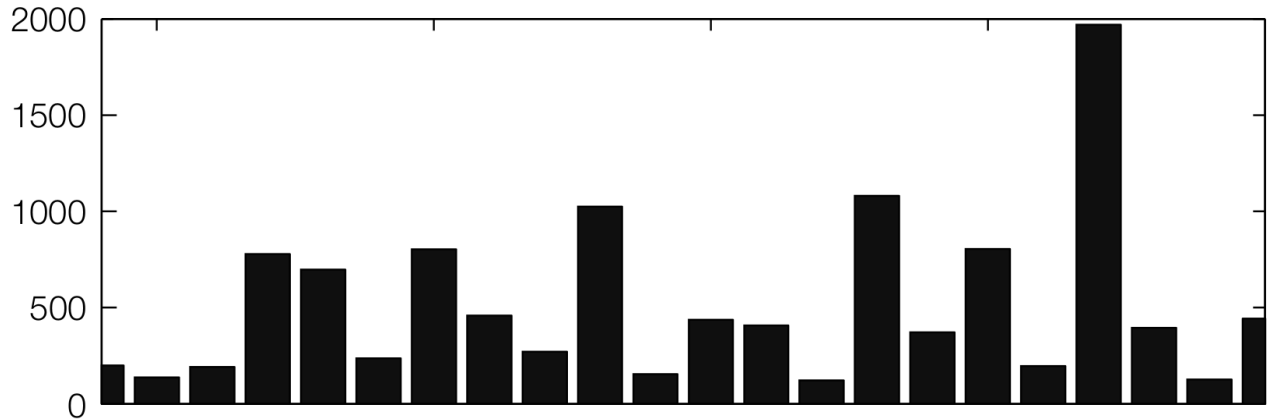
Using LSB.

同一个桶内进行交换. [相邻2个像素为一个桶内]

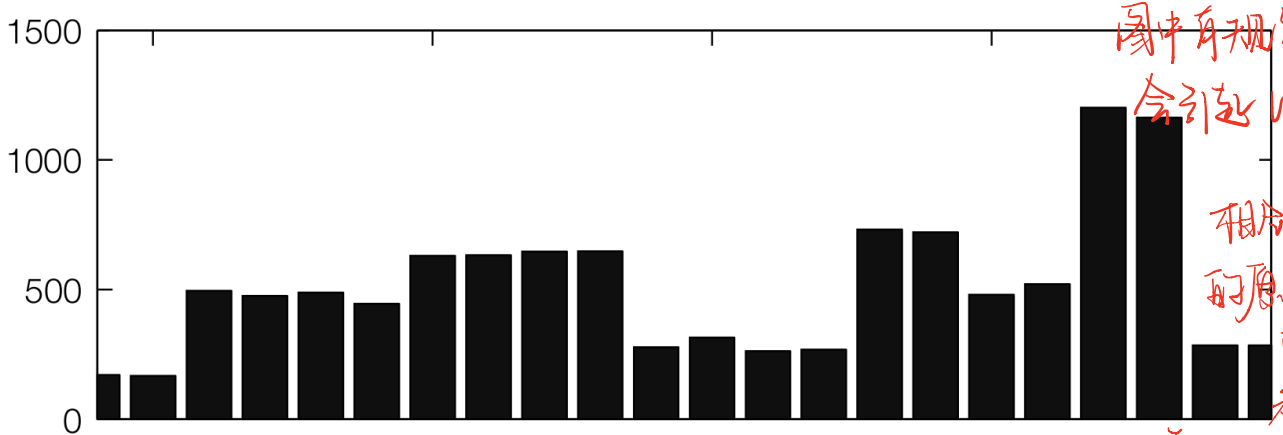
pixel values can be divided into disjoint pairs of values

- $(2i, 2i+1)$
- $2i \to 2i+1 : 1,\ 2i+1 \to 2i : 0.$

A Comparison



normal picture



LSB picture

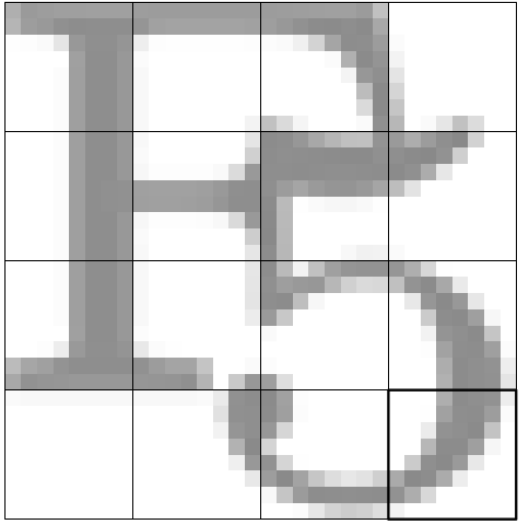
图中有规律
会引起 warden 注意
↑
相邻个数类似
的原因：加密水印
后，2i 和 2i+1
和 embed 的
基本上 i 和 0 个数相差：(m) 有关.

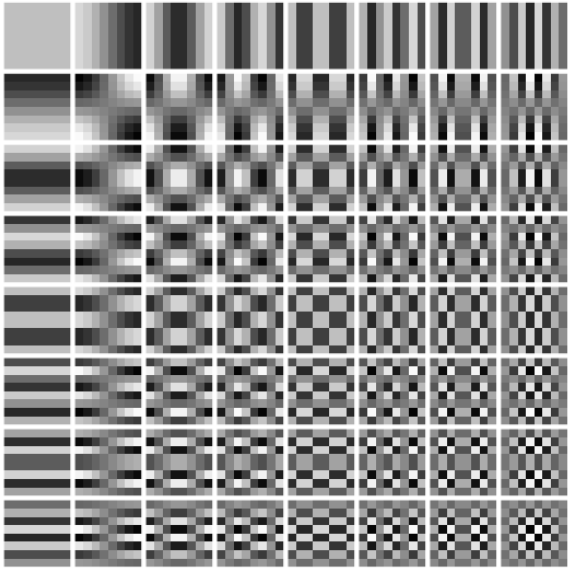
Practical Steganographic Methods

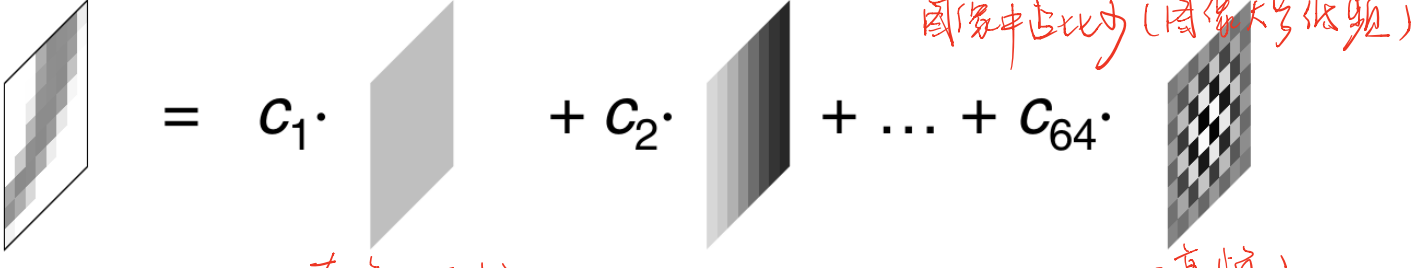
- OutGuess
- Masking Embedding as Natural Processing

DCT Coefficients

Discrete Cos transformation





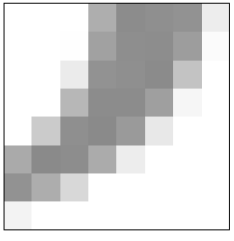


图像中占比少（图像大多低频）

直流（低频）

（高频）

DCT Compression 可以做压缩.

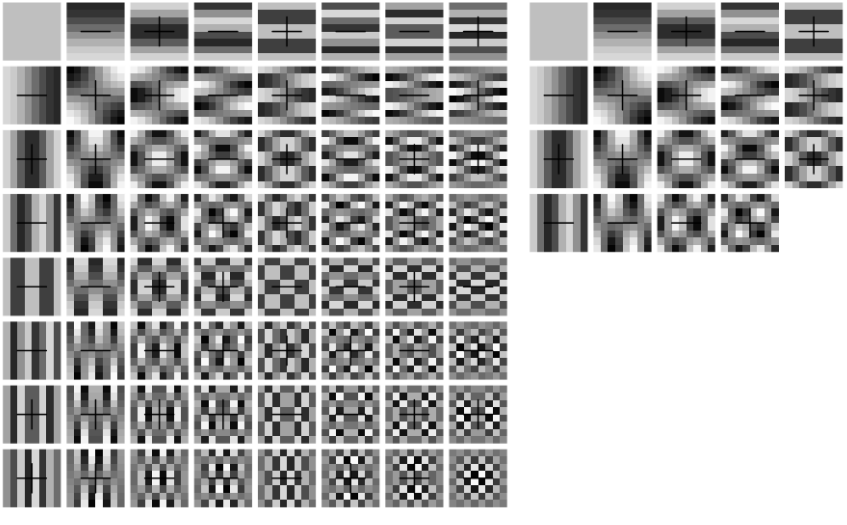


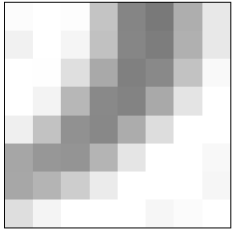
64 brightness values

只需要在少部分.

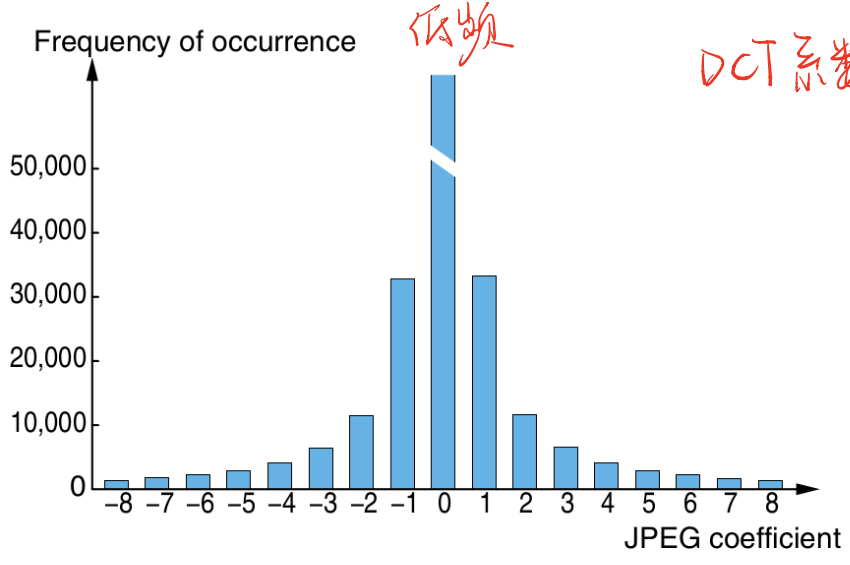
➽ 19 nonzero JPEG coefficients

存储这DCT系数.





DCT Characteristic Properties



$$P(X=1) > P(X=2) > P(X=3) > P(X=4)$$

$$P(X=1)-P(X=2) > P(X=2)-P(X=3) > P(X=3)-P(X=4)$$

OutGuess

- Preserving DCT Statistics
  - first pass: LSB along a pseudo-random walk
  - second pass: correct the coefficients to restore the histogram
- The maximum length that can be embedded
  - Ensuring that one will be able to make corrections
  - determined by the frequencies of the most unbalanced LSB pair.

For Simple Detection

embed 比例小取近高.

In a bin consists of a pair of values (U, L). In normal work, $U > L$. Let fraction $q \in [0, 1]$ of the bin is used to embed, how large q could be?

cover	U	L
unchanged	$U \cdot (1 - q)$	$L \cdot (1 - q)$
changed	$(U + L) \cdot \frac{q}{2}$	$(U + L) \cdot \frac{q}{2}$
sum	$U - (U - L) \cdot \frac{q}{2}$	$L + (U - L) \cdot \frac{q}{2}$

- Embedding: U decreases by $(U - L) \cdot \frac{q}{2}$.
- Restoring: at most $L \cdot (1 - q)$ can be turned to U.
- To make sure of recovering U:

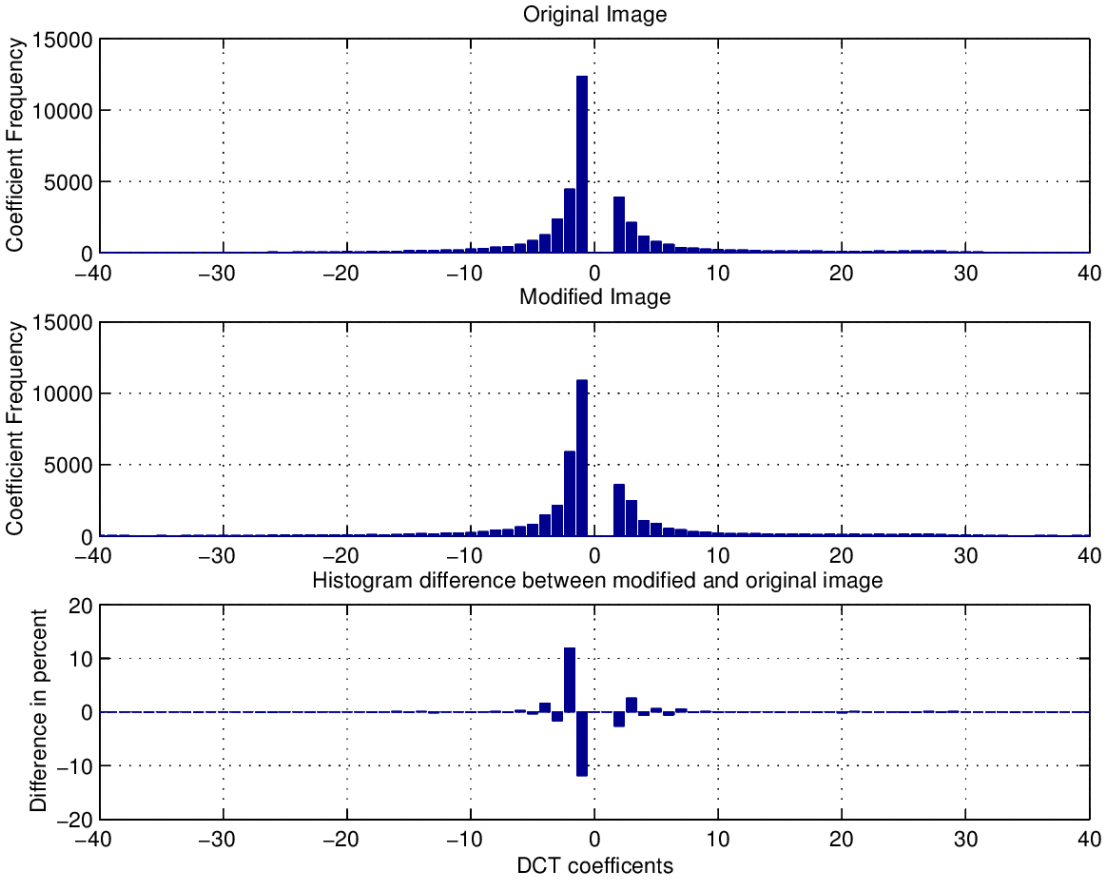
$$(U-L)\cdot\frac{q}{2} < L\cdot(1-q) \Rightarrow \frac{q}{2}(U+L) < L \Rightarrow q < \frac{2L}{U+L}.$$

Capacity

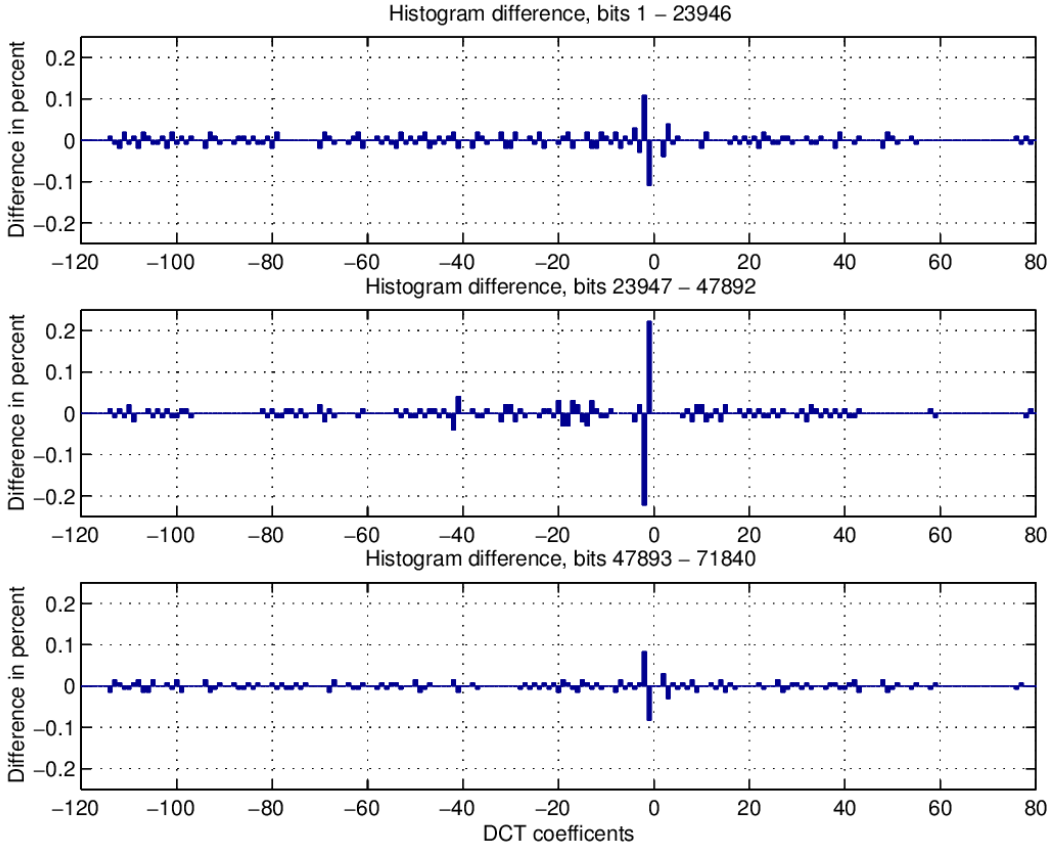
- Embedding capacity.
- Steganographic capacity.

Naive Embedding

0,1 不动（拐例高度区域）



More Advanced Method



Defending Against Statistical Steganalysis, 2011, 10th USENIX Security Symposium

Basic Idea

Each bin contains a lots of pixel pairs.

- Some of them for embedding.
- Some of them for correction.

Identical histogram

- One embedding goes with one correction.

Model-Based Steganography

数据压和校验压在一起.

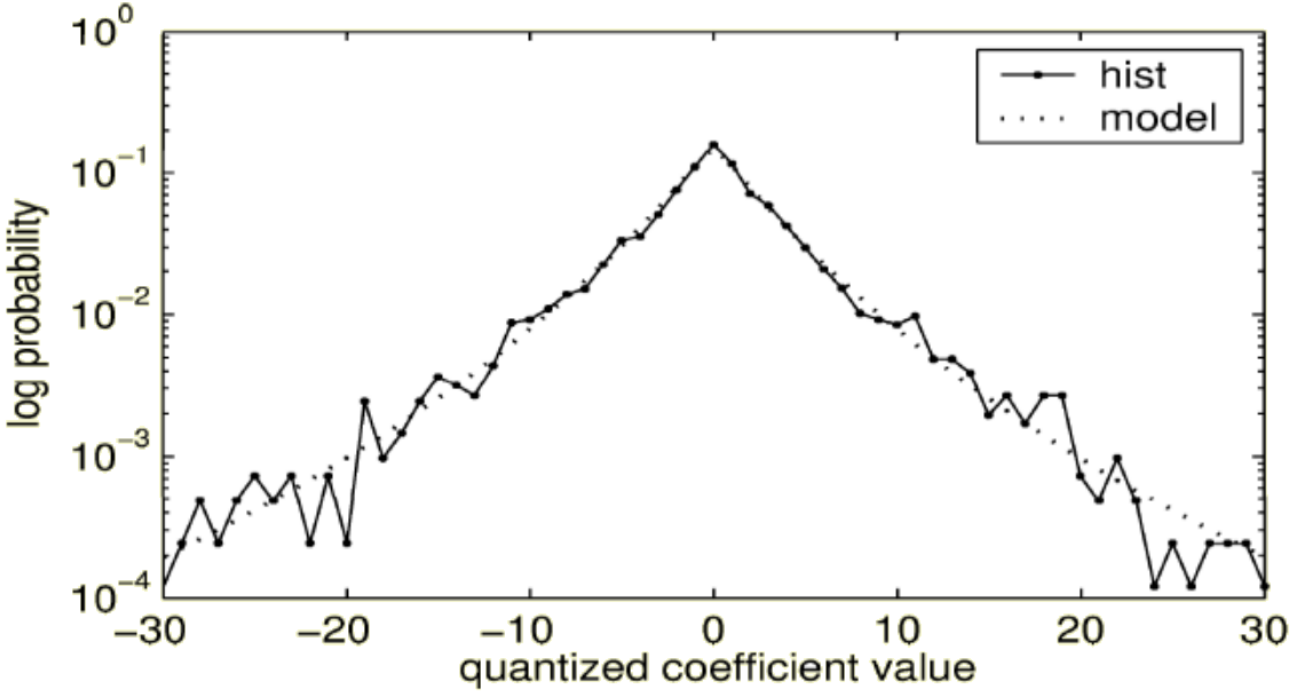
Generalized Cauchy model with probability density function (pdf)

- Generalized Cauchy distribution (GCD):

$$P(x) = \frac{p-1}{2s}\left|\frac{|x|}{s}+1\right|^{-p}.$$

- $p > 1, s > 0$ are the two parameters.

Illustration of GCD



Two-Class Pattern Classification

Two components in a cover work (c_{inv}, c_{emb}):

$$
\begin{aligned}
p_0 &= P(c_{emb} = 0 | c_{inv} = MSB_7(2i)) \\
&= \frac{T_c[2i]}{T_c[2i] + T_c[2i+1]} \\
&= 1 - P(c_{emb} = 1 | c_{inv} = MSB_7(2i)).
\end{aligned}
$$

The probability of $2i$ in the bin $(2i, 2i+1)$.

Arithmetic Decompress and Compress

Map a uniformly distributed bitstream to a new bitstream with specific distribution.
Presentation: Arithmetic Coding

- `http://en.wikipedia.org/wiki/Arithmetic_coding`
- `http://www.cs.cmu.edu/~aarti/Class/10704/Intro_Arith_coding.pdf`

Reverse Compression

- In embedding:

  uniformly distributed bitstream

  $\xrightarrow{\text{Decompress}}$

  GCD distributed bitstream

- In detection:

  GCD distributed bitstream

  $\xrightarrow{\text{Compress}}$

  uniformly distributed bitstream