Problem Description:

1) compute the steganographic capacity for odd-even based hiding in QDCT domain and SR based compensation (expressed through optimal hiding fraction and hiding rate)

2) generalize the method for various orders of co-occurrence statistics in QDCT domain

3) identifying which frequency components are more useful for detection after first order statistical restoration over the entire hiding band

(1) Computing Optimal Hiding Fraction & Rate for 1-D PMFs

Let $B_X(i)$ and $B_Y(i)$ denote the number of elements in the $i^{th}$ bins of $X$ and $H$, respectively.

Assuming almost equal number of 0s and 1s in the input message and uniformly distributed dither values and a hiding fraction of $\lambda$ for all the bins, we obtain

$$B_Y(i) = \frac{B_X(i)}{2} + \frac{2\delta}{4} B_X(i)$$

For perfect restoration, considering the $i^{th}$ bin, the hiding fraction $\lambda$ needs to satisfy:

$$B_Y(i) \leq B_X(i) \Rightarrow \lambda \leq \frac{B_X(i)}{B_X(i) + B_X(i+\delta)}$$

Defining $\lambda_i = \frac{B_X(i)}{B_X(i) + B_X(i+\delta)}$ and after hiding in QDCT coefficients $\in [-T, T]$, the effective hiding fraction $\lambda'(T)$, for a given threshold $T$, is given by:

$$\lambda'(T) = \min_{-T \leq x < T, \lambda_i < \lambda}$$

As the threshold $T \uparrow$, $G(T)$, the fraction of terms available for hiding $\uparrow$,

$$G(T) = \sum_{T < i < T} P_X(i)$$

where $P_X(i)$ is the pmf of $X$.

We select the threshold $T_{opt}$ for which the rate $R(T)$ is maximized (to produce $T_{opt}$).

$$R(T) = \lambda'(T) G(T)$$

$$T_{opt} = \operatorname{arg \max}_T R(T)$$

(2) Generalization for $n^{th}$ order co-occurrence statistics

A single bin now has $n$ elements, say $\{i_1, i_2, ..., i_n\}$. Perturbation of only $\pm 1$ is allowed in odd-even QDCT domain hiding.

$i_1$ component can be mapped to $i_1, i_1 \pm 1$ with probability $\frac{1}{2}$ and $\frac{1}{2}$, respectively. Thus, $\{i_1, i_2, ..., i_n\}$ can be mapped to $\{i_1 + \delta_1, i_2 + \delta_2, ..., i_n + \delta_n\}$, where $\delta_j \in \{-1, 0, 1\}, 1 \leq j \leq n$.

Assuming a hiding fraction of unity,

$$B_Y(i_1, i_2, ..., i_n) = \sum_{i_1 \leq i_1 \leq n} \sum_{i_2 \leq i_2 \leq n} \sum_{i_3 \leq i_3 \leq n} \ldots \sum_{i_n \leq i_n \leq n} f(i_1)f(i_2)f(i_3)\ldots f(i_n)$$

where $f(0) = \frac{1}{2}, f(1) = \frac{1}{2}, f(-1) = \frac{1}{2}$

$$\lambda(i_1, i_2, ..., i_n) = \frac{B_Y(i_1, i_2, ..., i_n)}{B_X(i_1, i_2, ..., i_n)}$$

$$X'(T) = \min_{-T \leq T \leq T} \{\lambda(i_1, i_2, ..., i_n) : \lambda(i_1, i_2, ..., i_n) > 0\}$$

(3) Total Compensation vs Individual Compensation

Let steganographer perform total compensation (restores first order pmf for the entire hiding band) and stegananalyst consider individual hiding band pmfs, for each of the first 19 AC QDCT frequency terms (zigzag scan order).

We expect pmf mismatches for individual bands but will there be a general trend to the mismatch for proper selection?

We use 4500 images for the detection experiments – half for training and half for testing, where training and testing sets have equal number of cover and stego images, and perform SVM based steganalysis.

**Statistical Restoration Setup**

Let hiding fraction $\lambda = |H|/|X|$ – as $\lambda \uparrow$, difference between pmf of $(H \cup C)$ and pmf of $X$ $\uparrow$.

As $\lambda \uparrow$, $|C| \downarrow$ and compensating power of $C$ $\downarrow$.

There exists a minimum value of $\lambda$, called $\lambda_{opt}$, such that for $\lambda > \lambda_{opt}$, $C$ can no longer compensate for PMF mismatches due to hiding.

The aim is to compute $\lambda_{opt}$ given $X$ for maximal hiding and maintaining statistical transparency.

Odd even based hiding framework to embed 1: $q=\text{round}(p+1 - \text{mod}(p+1, 2))$

To embed 0: $q=\text{round}(p+1 - \text{mod}(p+1, 2))$

where $p$, the original quantized DCT (QDCT) term, is mapped to $q$, $\delta$ is obtained from a dither sequence which takes values in $[-0.5, 0.5]$

### Steganographic capacity

For a given host, it is the maximal message length that can be embedded subject to

a) **Perceptual transparency**: ensured by choice of hiding band and perturbation of $\pm 1$ per QDCT term (odd-even based hiding)

b) **Statistical transparency**: using $\lambda = \lambda_{opt}$ in SR setup

It depends on the transform domain used for hiding, hiding method and image statistics compensated for

**Total compensation based hiding** is most detectable when QDCT(1,6) PMF is used for detection – this occurs due to the “very peaked” nature of the PMF of QDCT(1,6) near 0

$B_a(0) > B_a(1), B_a(-1), \Rightarrow \lambda_i = 0$ for $i = \pm 1$

$\Rightarrow$ during total compensation, we consistently hide much more data in QDCT(1,6) than permitted by $X'(T)$

$\Rightarrow$ better detection using QDCT(1,6) PMF

Comparing PMF differences, after SR, for individual and total compensation schemes, we

- **Orig**: original PMF of the particular QDCT band
- **Indiv**: PMF after hiding and individual compensation
- **Total**: PMF after hiding and total compensation

There is significant PMF mismatch in case of QDCT(1,6) (difference is 0.10 for QDCT(1,6) & 0.003 for QDCT(5,1))