universität
innsbruck

Information Hiding

Exercise Sheet 1

Summer Term 2025

20 March 2025

Preparation for the lab session on **27 March 2025**.

1 Processing an image

For the following task, use the image `10.png` from the BOSSBase database.

a) Increment all the pixels by 1. What PSNR do you expect? Measure the PSNR experimentally.

b) Apply gamma correction, $\gamma = 1.1$, and store the resulting image in PNG. Compare the image histograms. Plot the distribution of the quantization error.

c) Subsample the image by nearest neighbor. Mark the areas with aliasing. Employ a linear filter prior to the subsampling to suppress the aliasing.

d) Compute the 2D DCT spectrum of the image (over the whole image, not in 8×8 blocks like JPEG) and apply top-left cropping. What effect does it have when transferring it back to the spatial domain?

e) Read the DCT coefficients of a JPEG-compressed version of this image, for example by using the Python package `jpeglib` or by calling the standard `libjpeg` library directly. Compare the histograms of the DC (0, 0) and AC (11, 44) modes.

2 Making decisions

Data batch $\boldsymbol{X}$ with labels $\boldsymbol{y}$ is passed to a detector that produces predictions $\hat{\boldsymbol{y}}$.

$\boldsymbol{y}$	0	0	1	1	0	1	0	0	1
$\hat{\boldsymbol{y}}$	0.4	0.45	0.55	0.45	0.4	0.6	0.55	0.45	0.6

a) Build the confusion matrix for the decision threshold $\tau = 0.5$.

b) Calculate the accuracy, precision, and recall metrics.

c) Draw the ROC curve. Calculate ROC AUC, equal-error rate, and the probability of error P_E.

3 Bit by bit

Consider a cover element $x_i^{(0)}$, a stego element $x_i^{(1)}$ and a message bit m_j, stored in variables `x0`, `x1`, `m`.

a) Embed the message `01010011` using sequential LSBR into the following cover

$$x^{(0)} = (105, 105, 116, 98, 105, 104, 104, 107, 101, 114).$$

b) Write a line of C code that replaces the LSB of `x0` with `m` using

- arithmetic operators:
- bit masking:
- bit shifting:
- bitwise xor:

c) Simulate $\alpha = 0.4$ of LSBR into an image using `conseal`, and measure the empirical change rate $\hat{\beta}$, embedding rate $\hat{\alpha}$, and embedding efficiency $\hat{e}$.

d) Adapt the F5 estimator $\hat{\beta}$ from the lecture to LSB matching in the spatial domain. Why can such an attack not be used in practice?

4 Theoretically secure

In an ideal world, where continuous variables can be perfectly represented in the computer, let us have a normally distributed cover $\boldsymbol{x}^{(0)} \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$ with N elements. The embedding, $\boldsymbol{x}^{(1)} = \boldsymbol{x}^{(0)} + \boldsymbol{\delta}$, involves adding a noise $\boldsymbol{\delta} \overset{i.i.d.}{\sim} \mathcal{N}(0, \gamma\sigma^2)$, $\gamma \in \mathbb{R}^+$, $\boldsymbol{\delta} \perp \boldsymbol{x}$.

- Express the theoretical security for the embedding, parameterized by N and γ, using a log-likelihood ratio and a threshold $\log(\tau)$.
- Given $N = 1000$ samples, what is the maximal value of γ to achieve security above $\log(\tau) = 0.9$?
- Cross-check your previous answer with direct estimation of the LRT on simulated stego vectors.

5 Innocuous and despicable

The archive "`DCIM.zip`" contains files that were captured from a communication of Martin and his colleague Verena.

a) Analyze the images for LSBR. Which of them seem(s) to carry a steganographic payload?

b) Try to extract the message. It may require a pinch of detective work.