Digital Steganography
Least Significant Bit Manipulation of Image Files
Johnnie Thomas
Faculty Advisor: Jianbo Shi

ABSTRACT
The project resulted in an application for embedding and extracting information (text messages) in red/green/blue images by altering the least significant bits (LSBs) of pixels within the image. The application increases security by using a private key, distributed encoding, and relative encoding.

PRIVATE KEY
- The private key is 9 digits, $R_i, R_j, C_{i,j}, G_{i,j}, Q_{i,j}$ required by both the encoder and the decoder.
- $(C_{i,j}, R_{i,j})$ is the coordinate of the origin pixel in the encoding/decoding process. If this is outside the bounds of the image’s dimensions $(M \times N)$, tile the image will yield the origin pixel:

$$\frac{Q_{i,j}}{1000}$$

is the maximum percentage of an image’s pixels that are usable in the encoding/decoding process given the key.

DI STRIBUTED ENCODING
- The system determines $P_i$ and $P_j$, the 2 pixels that will carry the $i$-th character of the message.
- These pixels are distributed throughout the image. $\frac{1000}{Q_{i,j}}$ approximates the average separation between altered pixels in an image encoded with a message of maximal length for the given key.

RELATIVE ENCODING
- There are 4 encoding functions ($E_1, E_2, E_3, E_4$) and 4 decoding functions ($D_1, D_2, D_3, D_4$) for $E(x) \neq E_j(x), \forall j \in \{1, 4\}, k \in \{1, 4\}, x \in \{1, 64\}$, where $j \neq k$.

- The system chooses 1 of these encoding functions based on the color of $P_i$ to determine the 6-bit sequence, new_c, that will be encoded in $P_i$ and $P_j$ to represent the $i$-th character of the message:

$\text{new}_c = a^b \quad a = \{R,G,B\}$

- Original characters can be retrieved given the new image, the key, and the above decoding functions. This is to defend against steganalysis (detection) techniques that examine images for regions highly concentrated with possible tampering evidence.

ENCODING SUMMARY

The chart above explains the encoding process, beginning with $i = 1$ and continuing until the END state is reached or there are no more pixels to alter given a particular image and key combination (in other words, $Q_{i,j}/1000$ of the pixels in the image have been used). The decoding process follows a similar path to examine pixels and uses the decoding functions to recover the original characters of the message.

SAMPLE IMAGES
Below are 3 images encoded with the same message (The Declaration of Independence) using the same key (070417076). Shown are the original images, the encoded images, and histograms of the bits used to encode pixels in the red, green, and blue layers of each image. As a result of relative encoding, the encoding bit histograms are not identical (i.e. the images were not identically altered).

REMARKS
- This encoding works in the space of LSBs, which are not preserved during compression. Encoded images must be saved and transmitted in an uncompressed format, which can become cumbersome.
- There is an effort to develop methods for detecting manipulations of uncompressed images by examining the ratio of unique colors to total pixels in the image; however, no reliable methods currently exist.