

Haneen Ameen Zamzami

Steganalysis is described as "the art and science of detecting secret messages hidden using steganography" and focuses on detecting hidden files, images, and text. Steganalysis is a technique for detecting hidden information by analyzing data transmission, but it does not make any assumptions about the steganography algorithm used. Steganalysis is commonly used in computer forensics, electronic security, and internet monitoring of illegal activity. Steganalysis also aids in the improvement of steganographic method protection by identifying and finding weaknesses.

In this proposal, we propose a system for detecting steganographic images by extracting number of image features such as correlation, contrast, homogeneity, and energy. Extracting different image features is one of the best techniques in blind steganalysis. It makes the system more able to deal with many types of colored images regardless of their sizes and the steganography methods were used to hide the data.

Question/need:

The proposed system helps to investigate any illegal behavior as well as prevent any unwanted events from occurring because of hidden communication. The developed method of steganalysis is used to examine the possibility of hidden messages in images. Also to combat particular techniques of steganography, several steganalysis methods have been introduced. However, no particular detection method can accept all picture formats or detect all steganography types. Thus, there must be more steganalysis methods or techniques to cover a broad spectrum of image formats and colours.

Data Description:

Build a validation image dataset. Detection of some stego-images or use of some image steganography techniques to create stego-images for testing with the modified method.

I download the images from different resource Kaggle , pixabay and gettyimages. then I create my own database by extract the Gray Level Co-Occurrence Matrix (GLCM) properties of correlation, contrast, homogeneity, and energy .The database it contains more than 4000+ records with 8 features

features:

image name

Image size

Image format (JPG- BMP- TIFF)

Contrast of image

Homogeneity of image

Energy of image

Correlation of image

Catogry of Images (clean\stego)

```
%matplotlib inline
import pandas as pd
import matplotlib.pyplot as plt
images = df
images.head()
```

	Image Name	size	Energy	Contrast	Homogeneity	Correlation	format	catogary
0	00000000.jpg	8839	0.11872	0.12082	0.94235	0.98761	jpg	Clean
1	00000000_(2).jpg	10189	0.12359	0.32653	0.91042	0.96267	jpg	Clean
2	00000000_(3).jpg	8406	0.11269	0.41517	0.88805	0.93404	jpg	Clean
3	00000000_(4).jpg	6397	0.13714	0.24250	0.88940	0.96415	jpg	Clean
4	00000000_(5).jpg	5612	0.13951	0.36196	0.92156	0.94602	jpg	Clean

```
print(images.shape)
```

(4673, 8)

Tools:

- Jupyter notebook or colab
- matplotlib
- Pandas
- Numpy - sklearn

MVP Goal:

Use classification method

Spilt database to train and test