

Edith Cowan University

Research Online

Australian Digital Forensics Conference

Conferences, Symposia and Campus Events

2017

Building a dataset for image steganography

Chris Woolley

School of Science, Edith Cowan University, cwoolle1@our.ecu.edu.au

Ahmed Ibrahim

School of Science, Edith Cowan University, ahmed.ibrahim@ecu.edu.au

Peter Hannay

Security Research Institute, Edith Cowan University, p.hannay@ecu.edu.au

Follow this and additional works at: <https://ro.ecu.edu.au/adf>

 Part of the [Computer Sciences Commons](#)

Recommended Citation

Woolley, C., Ibrahim, A., & Hannay, P. (2017). Building a dataset for image steganography. DOI: <https://doi.org/10.4225/75/5a83a0541d284>

DOI: [10.4225/75/5a83a0541d284](https://doi.org/10.4225/75/5a83a0541d284)

Woolley, C., Ibrahim, A., & Hannay, P. (2017). Building a dataset for image steganography. *Paper presented in Valli, C. (Ed.). The Proceedings of 15th Australian Digital Forensics Conference 5-6 December 2017, Edith Cowan University, Perth, Australia.*

This Conference Proceeding is posted at Research Online.

<https://ro.ecu.edu.au/adf/170>

BUILDING A DATASET FOR IMAGE STEGANOGRAPHY

Chris Woolley¹, Ahmed Ibrahim², Peter Hannay²

²Security Research Institute, ¹School of Science, Edith Cowan University, Perth, Western Australia
cwoolle1@our.ecu.edu.au, ahmed.ibrahim@ecu.edu.au, p.hannay@ecu.edu.au

Abstract

Image steganography and steganalysis techniques discussed in the literature rely on using a dataset(s) created based on cover images obtained from the public domain, through the acquisition of images from Internet sources, or manually. This issue often leads to challenges in validating, benchmarking, and reproducing reported techniques in a consistent manner. It is our view that the steganography/steganalysis research community would benefit from the availability of common datasets, thus promoting transparency and academic integrity. In this research, we have considered four aspects: image acquisition, pre-processing, steganographic techniques, and embedding rate in building a dataset for image steganography.

Keywords: Dataset, Image Steganography, Steganalysis, Embedding Rate

INTRODUCTION

Common datasets are widely used in different domains such as Image Processing, Artificial Intelligence, Cyber Security, etc. Some popular examples include FVC2002 (Maio, Maltoni, Cappelli, Wayman, & Jain, 2002) used in biometrics, PhysioBank Databases (Goldberger, et al., 2000) consisting of physiological datasets, and UCI KDD datasets (Hettich and Bay, 1999) used in data mining. Common datasets that are publicly available allow researchers to validate, benchmark, and reproduce techniques previously reported or proposed, thus promoting transparency and integrity of academic research.

We have carefully chosen the title of this paper as a “dataset for image steganography” in order to scope our work to only images. Thus, we would like to define a steganographic dataset as follows:

“A steganographic dataset is a collection of media objects consisting of cover objects and corresponding steganographic objects, with variations in parameters related to the objects and steganographic techniques used, applied consistently across all objects”.

In our definition, objects may include digital file types such as text, image, audio, and video. Therefore, this paper only focuses on image steganography dataset. Additional terms such as cover objects, steganographic objects, and relevant parameters are discussed in detail in later sections of this paper.

To the best of our knowledge, there are no such datasets for image steganography. Thus datasets used in existing literature have been constructed typically by the authors using public domain images. However, we argue that using public domain images introduces the risk of the integrity of such images being compromised, as absence of steganographic content is not known a priori.

The next section provides a detailed background on concepts that influenced our decisions in building the dataset. This section is followed by the method, which outlines the experimental design and further details on activities that were carried out in each step of the experiment. The paper is then concluded with final remarks and avenues for future direction.

BACKGROUND

Steganography first emerged as a form of covert communication thousands of years ago employing techniques such as using invisible ink to hide a message in plain sight or masking secret messages within inconspicuous text (Pieprzyk, Hardjono, & Seberry, 2003). Different forms of secret communication have evolved through the centuries into two specific sciences, cryptography and steganography. Cryptography deals with rendering a secret message unreadable to anyone other than the intended recipient. In contrast, steganography hides the existence of the secret message within cover objects such as text, image, audio, or video files. An unsuspecting party would not be aware of the presence of a secret message even if they came to possess the file. However, if the presence of

the secret message were discovered, then the goal of steganography would be defeated. This form of discovery, attack, or detection, is formally known as steganalysis.

This section informs the reader with background information about factors in the literature that influenced the criteria and methodology for creating the dataset for image steganography. We identified four key factors as: image acquisition, pre-processing, steganographic techniques, and embedding capacity.

Image Acquisition

Three aspects of image acquisition that stood out from the literature were its source, type, and quantity.

Image acquisition

Three sources were predominantly used throughout the literature were:

1. public domain image datasets,
2. downloading images from public domain Internet sources, and
3. manually acquiring images.

The image datasets available in the public domain cited in the literature were NRCS (Pevny et al., 2010; Huang et al., 2010; Ker, 2005), Greenspun (Aycibas et al., 2005; Farid, 2002; Fridrich, 2004; Lyu & Farid, 2002), and ImageNet (Zeng et al., 2016) Freefoto (Lyu & Farid, 2004) and the use of stock photo or clip-art compilations (Huang, et al., 2010). However, Ker (2004) purchased the rights to the galleries used within his research.

In order to download images from Internet sources, Kharrazi et al. (2006) used web crawlers to scrape publicly available images. Others used established steganographic sources including the BOWS2 (Pevny et al., 2010) and BOSSBase (Kodovsky & Fridrich, 2013) databases.

According to Holotyak et al., (2005) and Pevny et al. (2010), both the original source and quality of the images can affect the effectiveness of steganalysis. Thus, some researchers developed their own image datasets, which allowed them to have control over the images used for their specific needs, for example, using multiple image capture sources to more closely reflect real-world application (Holotyak et al., 2005; Huang et al., 2010; Pevny et al., 2010).

Types of Image

The types of images used can be classified as either *natural* or *unnatural* images. Farid and Lyu (2003) defined natural images as photographs collected during standard or digital photography, allowing them to mimic real-world application of steganography. Within the same paper, they defined unnatural images as artificial images that have been created digitally. Additionally, Fridrich and Goljan (2002) discussed the suitability of types of images and proposed that cover images with low colour counts or unnatural images (e.g. digital art) should be avoided. This view is seemingly supported as there is an identified gap in the literature for images of these types.

Number of images

The number of images used varies drastically between the types of research conducted, with most literature exploring steganalysis techniques using quantities between 1,000 and 10,000 cover images (Aycibas, Kharrazi, Memon, and Sankur, 2005; Farid, 2002; Farid and Lyu, 2003; Fridrich, 2004; Holotyak, Fridrich & Voloshynovskyy, 2005; Huang, Shi & Huang, 2010; Ker, 2004; Ker, 2005; Kodovsky & Fridrich, 2013; Kodovsky, Pevny & Fridrich, 2010; Lyu & Farid, 2002; Pevny, Bas & Fridrich, 2010; Solanki, Sarkar & Manjunath, 2007). However, Zeng et al. (2016) argued that these numbers were not sufficient to represent real-world application and used 14,000,000 unique images of varying complexity from the ImageNet database instead. Other large quantities used in literature range from 40,000 (Lyu & Farid, 2014) up to one million (Kharrazi, Sencar, & Memon, 2006) unique cover images.

Conversely, literature that explores early steganalysis techniques (Fridrich and Goljan, 2001; Fridrich, Goljan, and Du, 2001) or new steganographic models are shown to use smaller cover image sets comprising of 10 or less images (Solanki, Sarkar & Manjunath, 2007; Zhang, Jiang, Zha, Zhang, & Zhao, 2013)

While these examples show that the number of images used vary between experiments, it is important to note that the literature does not discuss how the number of cover images may impact operational or acquisition costs, or if these factors have influenced the author's decisions.

Pre-processing

Pre-processing is used during the development of each image dataset to ensure consistency in the characteristics of the images used. This section explores how JPEG compression quality, resolution cropping, image complexity, and colour palettes are changed or observed during dataset creation.

JPEG Compression and Quality

One of the largest factors that affects the reliability of steganography within the JPEG domain is the original compression quality of the cover image. (Holotyak et al., 2005; Ker, 2005; Lyu & Farid, 2004; Pevny et al., 2010). When each JPEG image file is modified or saved, it undergoes *lossy* compression where it is compressed by stripping the image of imperceptible information, including fine details and colour gradients. In his book, Miano (1999) describes the four steps taken during JPEG compression as being:

1. downsampling Red Green Blue (RGB) colour palette to its luminance and chrominance components (YcbCr),
2. Discrete Cosine Transformation (DCT),
3. quantisation, and
4. Huffman Coding.

This process can make it easier for both a passive observer and steganalysis tools to detect differences within the image. Ker (2004) establishes the criticality of the JPEG compression process as he demonstrated higher susceptibility to steganalysis associated with higher compression rates. Holotyak, Fridrich, and Voloshynovskyy (2005) explained that this is because the compression process removed areas of high-frequency noise that would otherwise be used to efficiently and effectively embed the secret message. Because of this, experiments conducted on compressed JPEG images commonly used compression rates between 70% to 90% (Holotyak et al., 2005; Ker, 2005; Lyu & Farid, 2004; Pevny et al., 2010) and examples where compression quality were below 60% are harder to find (Ker, 2004; Ker, 2005; Kharrazi et al., 2006).

Image Cropping

Each dataset within the literature cropped the cover image file during pre-processing, with the two most common resolutions used being 512x512 (Holotyak et al., 2005; Ker 2004; Walia, Jain, and Naydeep, 2010) and 256x256 (Lyu and Farid, 2004) pixels. While other resolutions were uncommon, they did include 640x418 (Ker, 2005), 640x400 (Farid, 2002; Ker, 2005), 780x540 (Fridrich, 2004) 900x600 (Ker, 2004), 384x256 (Fridrich & Goljan, 2002), 1024x744 (Fridrich & Goljan, 2002), and 2100x1500 (Pevny et al., 2010) pixels.

As the resolutions used increase, so does the file size of the resulting image. One could intuitively assume that larger files can be used to embed larger payloads. However, Ker (2004) has shown that the increase in acceptable payload size and resolution are not proportional.

While not typically discussed within the literature, it is important to note that the cropping of images is likely due to the increased computational time and storage requirements that would be associated with larger images.

Image Complexity

While uncommon, some researchers (Liu, Sung, Xu, & Ribeiro, 2006; Lyu & Farid, 2004; Fridrich, 2004) took advantage of the cropping process to ensure that the areas used were consistent regarding image complexity and steganographic performance.

Liu, Sung, Xu, and Ribeiro (2006) also demonstrated that image complexity could affect the performance of steganalysis by specifically cropping both coloured and grayscale datasets and ranking the resulting images before performing steganalysis. This result showed that steganalysis was generally easier in images with lower complexity and that standard colour images and grayscale images with low complexity were comparable for most steganalysis techniques.

Colour Palettes

Using grayscale cover images is preferred for steganography as they are harder to detect during steganalysis compared to its colour counterparts (Liu, Sung, Chen, and Xu, 2008). According to Holotyak, Fridrich, and

Voloshynovskyy (2005), this is because it is more difficult to detect the steganography in grayscale images as the same inter colour channel relationships found in colour images do not exist in grayscale images. This is the preferred view in most literature with notable exceptions such as Fridrich (1999), Fridrich and Goljan (2002), Liu, Sung, Xu, and Ribeiro (2006), Lyu and Farid (2004), Rabie (2015), Wu and Noonan (2012) and Yu, Chang, and Lin (2007).

However, it is important to note that while grayscale is preferred in literature, real-world steganographic applications would include colour images and there is a lack of literature addressing colour images. Out of the 29 datasets examined, only the 7 papers identified above discussed or compared steganography in colour images.

Steganographic Techniques

Among the literature review for this paper, 29 discussed separate steganographic techniques that were used to build our dataset and their occurrence frequency has been listed Table 1, which clearly indicates the popularity of the individual techniques.

Table 1: Popularity of Steganographic Techniques

Steganographic Techniques	Frequency
LSB	10
Outguess	9
F5	7
Jsteg, Un-named	4
Ezstego, YASS, Steghide	3
Model-Based, s-tools, Hide4PGP, Steganos, MME3, PQ, NsF5	2
MME2, MSS, HUGO, LSBR, BCHopt, Edged-based, MB1, PQt, Pqe, Lqsteg, J-UNIWARD, UERD, UED, HQIH, JPHide, MM2, MM3	1

It is important to note that while this data shows that most sources used multiple algorithms or tools to create their dataset, when they did so it was often to compare algorithms that shared similar properties. Additionally, it was not uncommon for a single source to use a single tool (Holotyak, Fridrich & Voloshynovskyy, 2005; Ker, 2004; Ker, 2005; Kodovsky, Pevny & Fridrich, 2010; Rabie, 2015; Yu, Chang & Lin, 2007; Zhang et al., 2013)

Embedding Rate

The literature shows that the approach used to embed the payloads often varied and explored multiple variables for determining the capacity used. Common methods of embedding data involved using fixed payloads based on the size of the image (Aycibas et al., 2005; Farid, 2002; Kodovsky & Fridrich, 2013; Liu et al., 2008) or by embedding the information based on the Bits Per Non-Zero DCT Coefficient (BPC) of the images (Fridrich, 2004). In most cases, the payload size when using these methods was a multiple of 5 and ranged between 0.05 and 1.00. Notable exceptions to this were Ker (2004), who used a rate of 0.03 and Zhang et al. (2013) who included rates of 0.43, 0.61 and 0.83.

Other methods included embedded payload images with set resolutions instead of random data to create their steganographic images (Farid, 2002; Fridrich & Goljan, 2002; Lyu & Farid, 2002; Lyu & Farid, 2004; Monoharan et al., 2015; Rabie, 2015; Wu & Noonan, 2012; Yu et al., 2007). In these cases, the secondary image resolutions ranged from 32x32 to 256x256 pixels.

METHOD

As outlined in the previous section, the methodology for this research was structured based on the four aspects of image acquisition, pre-processing, steganographic techniques, and embedding capacity. Additionally, we also included a step to validate our processes by included hashing, logging, and validation of the dataset. A conceptual design of the experimental procedure is illustrated in Figure 1.

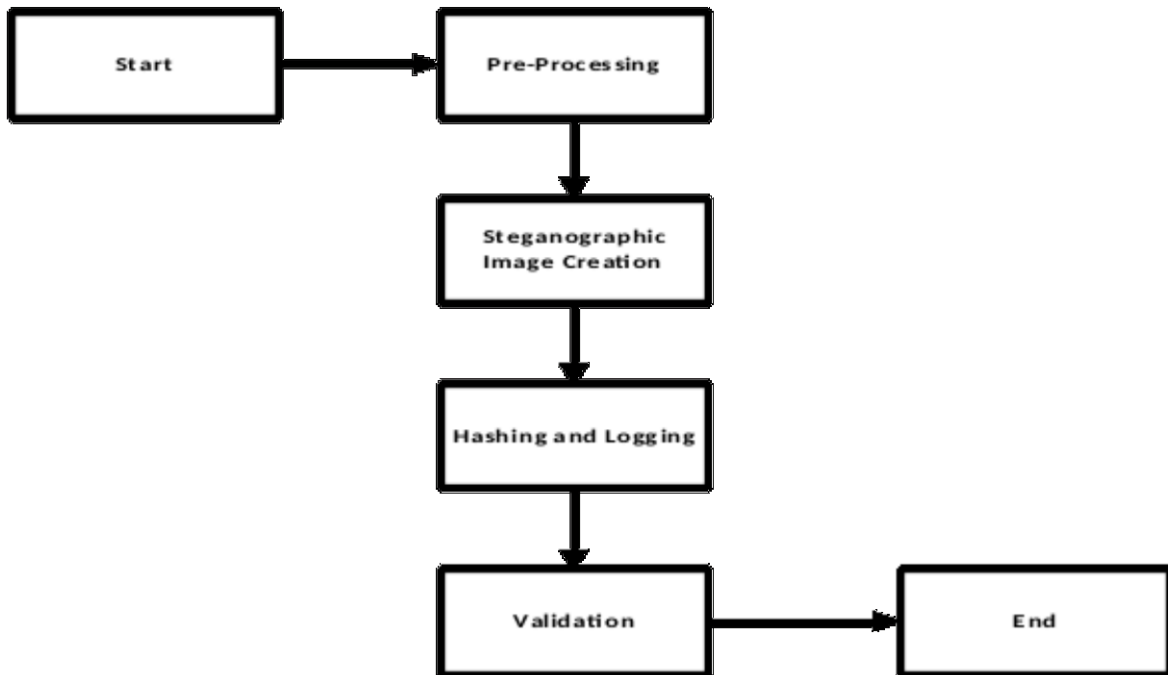


Figure 1: Conceptual Design of Experimental Procedure

Image acquisition

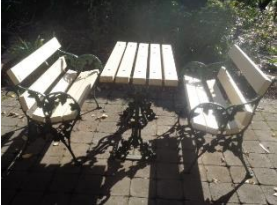





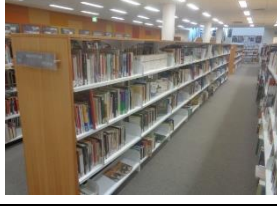
For the purpose of this research, the 1,760 images were acquired using a SONY DSC-W830 Cybershot camera (using base settings) for a duration of two months across a range of locations including public art installations, legal graffiti and street art zones, and public parks and buildings. Out of the final 1,000 images, 844 were taken in outdoor locations while 166 were taken inside public buildings. Further details of the image characteristics are listed in Table 2 while Table 3 provides visual examples of the types of images chosen. Lastly, during the image acquisition process, 760 images were discarded during the collection phase for reasons that included:

1. **Incorrect Settings:** leading to higher resolution and image sizes that would be unsustainable over the course of the cover image and its resulting steganographic image sets.
2. **Blurry/unusable images:** which would be unable to serve as an image for this purpose.
3. **Photography policies:** for public places such as the Australian War Memorial (Australian War Memorial, 2009) that we felt would not be compatible with the restrictions placed by the ethics committee approval.

Table 2: Cover Image Characteristics

Dimensions	640x480
Width	640 pixels
Height	480 pixels
Horizontal resolution	350 dpi
Vertical resolution	350 dpi
Bit depth	24
Compression	Uncompressed
Resolution unit	2
Colour representation	sRGB
Compressed bits/pixel	4

Table 3: Cover Image Examples

Cover Images	Sample Images		
Outdoor Settings			
Public Art Installations			
Legal Graffiti Zones			
Indoor Settings			

Pre-Processing

There were three main decisions to be made during pre-processing. These were:

1. What cropping and offset would be used?
2. What colour model to use?
3. What image types would be used?

Cropping and Offsetting

As the literature review showed, the two most common resolutions used for cropping steganographic cover images were 256x256 and 512x512 pixels. As the original cover images were only 640x480, we made the decision to crop the cover images to 256x256 pixels, which offers the added benefit of lower processing time due to fewer pixels. Furthermore, since all images were acquired manually, we did not see the need to offset during the cropping process.

Colour Palettes

As the literature identified a gap in the literature regarding colour-based steganographic datasets, we decided to use the coloured versions of the cover images wherever possible. It is important to note though, that some of the steganographic algorithms used required the use of grayscale images or converted the coloured cover images to grayscale during the embedding process.

As the processed cover image set does contain a set of grayscale images for the JPEG format, future experiments could use these files to create a compare between colour and grayscale images for JPEG steganography, which would assist with filling other identified gaps within the literature.

Image File Types

The image file types used during the creation of this dataset were JPEG and PGM. Other popular image file types such as TIFF, GIF, and BMP were not used in this dataset as they were not relevant to the steganographic techniques used. Table 4 shows the final properties for each set of cover images used.

Table 4: Post-Process Characteristics by Image Type

Property	Colour JPEG	Grayscale JPEG	Colour PGM
Dimensions	256x256	256x256	256x256
Width	256 pixels	256 pixels	256 pixels
Height	256 pixels	256 pixels	256 pixels
Horizontal resolution	350 dpi	350 dpi	350 dpi
Vertical resolution	350 dpi	350 dpi	350 dpi
Bit depth	24	8	24
Compression	Uncompressed	Uncompressed	Uncompressed
Resolution unit	2	2	2
Colour representation	sRGB	sRGB	sRGB
Compressed bits/pixel	4	4	4

The pre-processing operation is shown in Figure 2.

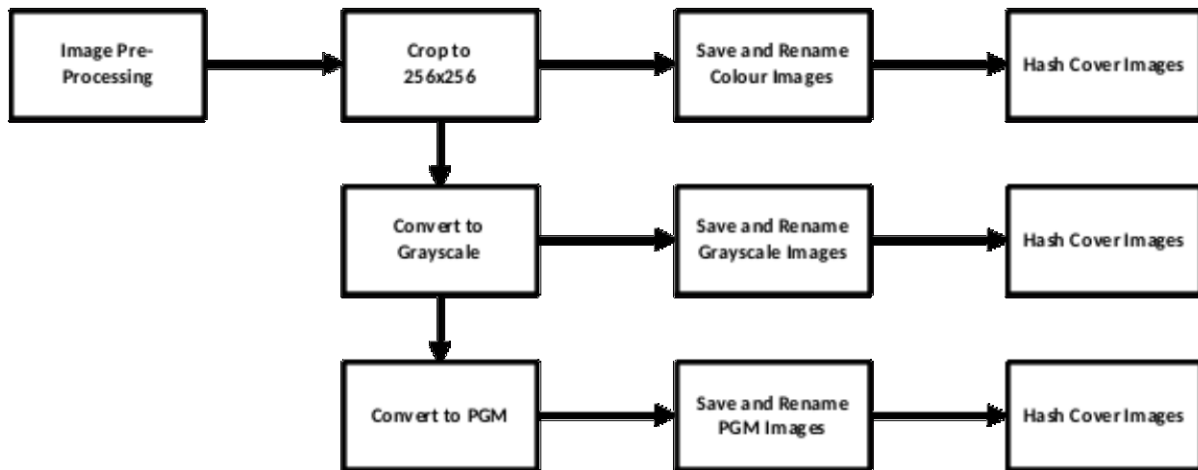


Figure 2: Pre-Processing Operation

Steganographic Techniques

The details of the steganographic techniques used during this research are summarised in Table 5 and Figure 3 shows the steganographic embedding process followed.

Table 5: Steganographic Techniques for Building the Dataset

Author	Algorithm or Tool	Domain	Input file type
Provos, N. (2001)	Outguess 0.2	JPEG	JPG
Westfield, A. (2001)	F5	JPEG	JPG
Hetzl, S. (2003)	Steghide 0.5.1	JPEG	JPG
DDE Lab. (2013b)	J-UNIWARD	JPEG	JPG
DDE Lab. (2013c)	S-UNIWARD	Spatial	PGM
DDE Lab. (2014)	SI-UNIWARD	Side-Informed	PGM
DDE Lab. (2013a)	nsF5	JPEG	JPG
DDE Lab. (2012a)	HUGO	Spatial	PGM
DDE Lab. (2012b)	WOW	Spatial	PGM
Kodovsky, J. (2007)	PQ	Spatial	JPG

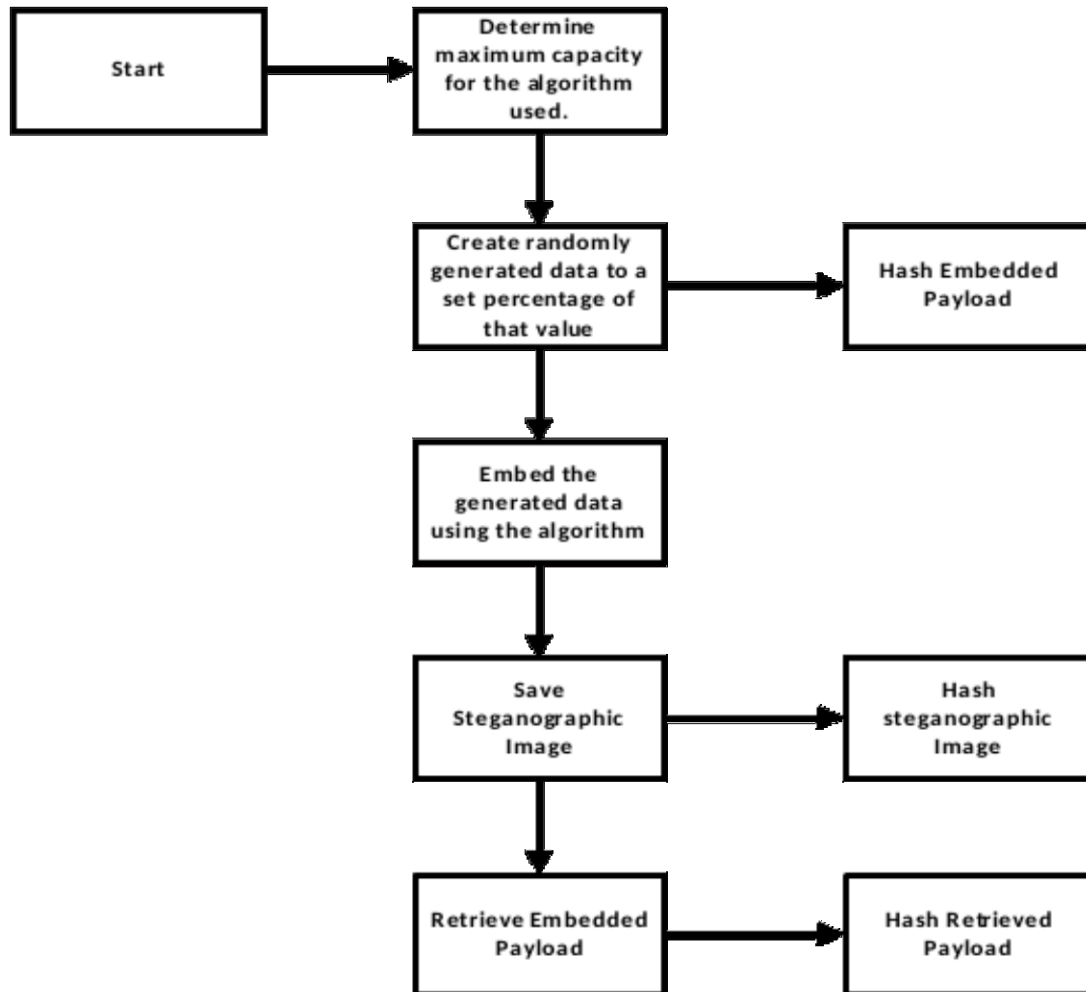


Figure 3: Steganographic Embedding Process

Embedding Rate

The embedding rate originally used within the dataset is 0.20 bits per available capacity (BPAC) for each cover image. As the BPAC will vary between each algorithm, this number must be calculated independently for each tool.

For Outguess and F5, this required embedding a small test file into each image to pull the estimated capacity from the tool's output (i.e., *-verbose* parameter). For Steghide, this required first viewing the maximum capacity (i.e., using *-info* parameter). Once these expected capacities were known, we were then able to calculate the required size of the test file based on the embedding rate selected.

For J-UNIWARD, S-UNIWARD, SI-UNIWARD, nsF5, HUGO, WOW, and PQ, the maximum embedding capacity is automatically calculated during the embedding process, and the only step required was using a variable to call the required embedding rate when executing these algorithms from the command line.

Verification and Dataset

Throughout the course of the experiment, various events were logged (date, time, embedding rate, and the path to artefact) and all individual artefacts created were hashed (MD5) for validation purpose. This allowed us to verify that the Cover images and Steganographic images were not the same after the embedding process and confirm that the exact payload could be retrieved when the tools used possessed this capability. At the end of the experiment, the final dataset we curated consisted of 14,000 images in total, which includes the original cover images, pre-processed cover images, and the final Steganographic images.

CONCLUSION AND FUTURE DIRECTION

This paper has presented the process and criteria we used to build a dataset for image steganography using. During our research, we discovered that there were different approaches used to construct the datasets used in literature in relation to steganography and steganalysis.

In our approach, we considered image acquisition, pre-processing, steganographic techniques, and embedding rate. During image acquisition, we chose to acquire manually using a digital camera, and used pre-processing to further format images to JPG and PNG, and also to crop the images to a specific resolution. Finally, we used multiple steganographic techniques and selected embedding rates determined consistent with each steganographic technique used.

One of the limitations we encountered was during the image acquisition process. Due to onsite photography policy, several images captured had to be excluded from the dataset. As a result, only a small portion of the dataset (16%) represents indoor images. We feel that there has to be a balance between indoor and outdoor images at least, in order to have a true representation of real-world applications.

REFERENCES

- Australian War Memorial. (2009). *Factsheet – Onsite Photography*. Retrieved from <https://www.awm.gov.au/sites/default/files/media/factsheet%20-%20onsite%20photography%20feb%2009.pdf>
- Avcibaş, İ., Kharrazi, M., Memon, N., & Sankur, B. (2005). Image steganalysis with binary similarity measures. *EURASIP Journal on Advances in Signal Processing*, 2005(17), 679350.
- DDE Lab, Birmingham University. (2012a). HUGO bounding dist. [Computer Software]. Retrieved from http://dde.binghamton.edu/download/stego_algorithms/
- DDE Lab, Birmingham University. (2012b). WOW. [Computer Software]. Retrieved from http://dde.binghamton.edu/download/stego_algorithms/
- DDE Lab, Birmingham University. (2013a). nsF5. [Computer Software]. Retrieved from <http://dde.binghamton.edu/download/nsf5simulator>
- DDE Lab, Birmingham University. (2013b). S-UNIWARD. [Computer Software]. Retrieved from http://dde.binghamton.edu/download/stego_algorithms/
- DDE Lab, Birmingham University. (2013c). J-UNIWARD. [Computer Software]. Retrieved from http://dde.binghamton.edu/download/stego_algorithms/
- DDE Lab, Birmingham University. (2014). SI-UNIWARD. [Computer Software]. Retrieved from http://dde.binghamton.edu/download/stego_algorithms/
- Farid, H. (2002). Detecting hidden messages using higher-order statistical models. In *Image Processing. 2002. Proceedings. 2002 International Conference on* (Vol. 2, pp. II-II). IEEE.
- Farid, H., & Lyu, S. (2003, June). Higher-order wavelet statistics and their application to digital forensics. In *Computer Vision and Pattern Recognition Workshop, 2003. CVPRW'03. Conference on* (Vol. 8, pp. 94-94). IEEE.
- Fridrich, J. (1999, April). A new steganographic method for palette-based images. In *PICS* (pp. 285-289).
- Fridrich, J. (2004, May). Feature-based steganalysis for JPEG images and its implications for future design of steganographic schemes. In *International Workshop on Information Hiding* (pp. 67-81). Springer Berlin Heidelberg.
- Fridrich, J., & Goljan, M. (2002). Practical steganalysis of digital images: State of the art. In *Electronic Imaging 2002* (pp. 1-13). International Society for Optics and Photonics.
- Fridrich, J., Goljan, M., & Du, R. (2001). Detecting LSB steganography in color, and gray-scale images. *IEEE multimedia*, 8(4), 22-28.
- Goldberger, A. L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., Mietus, J. E., Moody, G. B., Peng, C. K., Stanley, H. E. (2000). Physiobank, physiotoolkit, and physionet. *Circulation*, 101(23), e215-e220.
- Hettich, S. and Bay, S. D. (1999). The UCI KDD Archive [<http://kdd.ics.uci.edu>]. Irvine, CA: University of California, Department of Information and Computer Science
- Hetzl, S. (2003). Steghide 0.5.1. [Computer Software]. Retrieved from 'sudo apt-get install steghide'
- Holotyak, T., Fridrich, J., & Voloshynovskyy, S. (2005). Blind statistical steganalysis of additive steganography using wavelet higher order statistics.
- Huang, F., Shi, Y. Q., & Huang, J. (2010). New JPEG steganographic scheme with high security performance. In *International Workshop on Digital Watermarking* (pp. 189-201). Springer Berlin Heidelberg.
- Ker, A. D. (2004). Improved detection of LSB steganography in grayscale images. In *International workshop on information hiding* (pp. 97-115). Springer Berlin Heidelberg.

- Ker, A. D. (2005). Steganalysis of LSB matching in grayscale images. *IEEE signal processing letters*, 12(6), 441-444.
- Kharrazi, M., Sencar, H. T., & Memon, N. (2006). Performance study of common image steganography and steganalysis techniques. *Journal of Electronic Imaging*, 15(4), 041104-041104.
- Kodovsky, J. (2007). PQ. [Computer Software]. Retrieved from http://dde.binghamton.edu/download/stego_algorithms/
- Kodovský, J., & Fridrich, J. (2013, March). Quantitative steganalysis using rich models. In *IS&T/SPIE Electronic Imaging* (pp. 86650O-86650O). International Society for Optics and Photonics.
- Kodovský, J., Pevný, T., & Fridrich, J. (2010, February). Modern steganalysis can detect YASS. In *IS&T/SPIE Electronic Imaging* (pp. 754102-754102). International Society for Optics and Photonics.
- Liu, Q., Sung, A. H., Chen, Z., & Xu, J. (2008). Feature mining and pattern classification for steganalysis of LSB matching steganography in grayscale images. *Pattern Recognition*, 41(1), 56-66.
- Liu, Q., Sung, A. H., Xu, J., & Ribeiro, B. M. (2006, August). Image complexity and feature extraction for steganalysis of LSB matching steganography. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on* (Vol. 2, pp. 267-270). IEEE.
- Lyu, S., & Farid, H. (2002, October). Detecting hidden messages using higher-order statistics and support vector machines. In *International Workshop on Information Hiding* (pp. 340-354). Springer Berlin Heidelberg.
- Lyu, S., & Farid, H. (2004, June). Steganalysis using color wavelet statistics and one-class support vector machines. In *Electronic Imaging 2004* (pp. 35-45). International Society for Optics and Photonics.
- Maio, D., Maltoni, D., Cappelli, R., Wayman, J. L., & Jain, A. K. (2002). FVC2002: Second fingerprint verification competition. In *Pattern recognition, 2002. Proceedings. 16th international conference on* (Vol. 3, pp. 811-814). IEEE.
- Mathworks. (2017). MATLAB R2017a. [Computer Software]. Retrieved from https://au.mathworks.com/products.html?s_tid=gn_ps
- Miano, J. (1999). *Compressed image file formats: Jpeg, png, gif, xbm, bmp*. Addison-Wesley Professional.
- Pevny, T., Bas, P., & Fridrich, J. (2010). Steganalysis by subtractive pixel adjacency matrix. *IEEE Transactions on Information Forensics and Security*, 5(2), 215-224.
- Pieprzyk, J., Hardjono, T., & Seberry, J. (2003). *Fundamentals of computer security*. Berlin Heidelberg: Springer Verlag.
- Provos, N. (2001). Outguess 0.2 [computer software]. Retrieved from "sudo apt-get install outguess"
- Rabie, T. (2015). Lossless quality steganographic color image compression. *Int J Adv Comput Sci Appl*, 6(4), 114-123.
- Solanki, K., Sarkar, A., & Manjunath, B. S. (2007, June). YASS: Yet another steganographic scheme that resists blind steganalysis. In *International Workshop on Information Hiding* (pp. 16-31). Springer Berlin Heidelberg.
- Walia, E., Jain, P., & Navdeep, N. (2010). An analysis of LSB & DCT based steganography. *Global Journal of Computer Science and Technology*, 10(1).
- Westfield, A. (2001). F5 [computer software]. Retrieved from http://dde.binghamton.edu/download/stego_algorithms/
- Wu, Y., & Noonan, J. P. (2012). Image Steganography Scheme using Chaos and Fractals with the Wavelet Transform. *International Journal of Innovation, Management and Technology*, 3(3), 285.
- Yu, Y. H., Chang, C. C., & Lin, I. C. (2007). A new steganographic method for color and grayscale image hiding. *Computer Vision and Image Understanding*, 107(3), 183-194.
- Zeng, J., Tan, S., Li, B., & Huang, J. (2016). Large-scale JPEG steganalysis using hybrid deep-learning framework. *arXiv preprint arXiv:1611.03233*.
- Zhang, Y., Jiang, J., Zha, Y., Zhang, H., & Zhao, S. (2013). Research on embedding capacity and efficiency of information hiding based on digital images. *International Journal of Intelligence Science*, 3(02), 77.