Texture based steganalysis: results for color images

Patricia Lafferty, Farid Ahmed
The Catholic University of America, 620 Michigan Avenue, NE, Washington, DC 20064

ABSTRACT

Steganographic and watermarking information inserted into a color image file, regardless of embedding algorithm, causes disturbances in the relationships between neighboring pixels. A method for steganalysis utilizing the local binary pattern (LBP) texture operator to examine the pixel texture patterns within neighborhoods across the color planes is presented. Providing the outputs of this simple algorithm to an artificial neural net capable of supervised learning results in the creation of a surprisingly reliable predictor of steganographic content, even with relatively small amounts of embedded data. Other tools for identifying images with steganographic content have been developed by forming a neural network input vector comprised of image statistics that respond to particular side effects of specific embedding algorithms. The neural net in our experiment is trained with general texture related statistics from clean images and images modified using only one embedding algorithm, and is able to correctly discriminate clean images from images altered by data embedded by one of various different watermarking and steganographic algorithms. Algorithms tested include various steganographic and watermarking programs and include spatial and transform domain image hiding techniques. The interesting result is that clean color images can be reliably distinguished from steganographically altered images based on texture alone, regardless of the embedding algorithm.

Keywords: Watermarking, Steganography, Steganalysis, LBP, Local Binary Pattern, Texture Analysis, Neural Network

1. INTRODUCTION

1.1. Background

While several different genres of steganographic algorithms exist, all of them strive to hide from the casual observer the presence of data within some form of media cover file. If an observer recognizes anomalies within a steganographic media file, at that moment of realization the steganography fails to meet its purpose. Originators of steganographic techniques aim to shield their embedding techniques from detection by incorporating mechanisms into their hiding algorithms to avoid displaying one of the multitudes of characteristics researchers have identified to distinguish steganographic media from clean cover media.

For each new embedding algorithm, researchers have identified characteristic side effects that can be used to statistically identify media modified by that particular algorithm. Iteratively searching for any one of the many identified steganographic side effects in the media can not only help identify suspect media, but provide clues about the embedding technique employed. For known steganographic algorithms the test-and-see technique works fine.

Because of the nature of the field, steganographic algorithms are constantly enhanced to evade proposed detection algorithms. With the ever-multiplying number of programs and techniques available to embed data within media, it would be both wise and useful to determine a general detection method. We recognize that probably no one detection algorithm can perfectly detect every hidden message, but, undaunted, our goal is to make an attempt at general steganalysis by approaching the problem from a pattern recognition perspective.

This paper describes how to use the local binary pattern (LBP) operator to calculate statistics capturing the correlation between neighboring pixels, pixel neighborhoods, and across color planes to examine a color image for embedded data.

Mathematics of Data/Image Coding, Compression, and Encryption VII,
with Applications, edited by Mark S. Schmalz, Proc. of SPIE Vol. 5561
(SPIE, Bellingham, WA, 2004) 0277-786X/04/$15 · doi: 10.1117/12.559896

1.2. Steganalysis

Beginning in the middle to late 1990's, research efforts were initially put forth in the area of steganalysis, or the detection of steganography. By definition, the steganography has failed to meet its purpose if the presence of the data within the media is detectable. Steganalysis can be approached from two perspectives, each with a distinct goal. Passive steganalysis involves detection only. The steganalysis process ends when there is an answer to the question, "Does this media harbor steganographic data?" In the case of active steganalysis, the process is complete only after the hidden data is removed, destroyed, or strategically altered to render it useless. The process described in this paper involves only the determination of whether an image is innocent or steganographic media.

While the embedding techniques used in watermarking and steganography are fairly similar, the goals of each of the two areas are clearly distinct. Because of the limited space in which to place the data within the carrier media, there are tradeoffs that must be made in order to best suit the purpose of the embedding. Steganographic programs usually favor imperceptibility over robustness, whereas watermarking applications require robustness at the expense of imperceptibility or transmitted data quantity. Embedded data is a noise within the media. The more data placed inside cover media, the noisier the media file appears. The weaker the steganographic signal within the cover, the less robust it will probably to be manipulations, but the data may be almost imperceptible. With this background, the goal of steganographic utilities is primarily to hide data within a host imperceptibly. This may involve limiting the amount of data and changing the values of the pixels or coefficients in a neighborhood around the altered pixel to make the embedding more inconspicuous.

Because data located within the signal of a visual media file can be viewed as noise, there should be noticeable characteristics in the image texture that give away this fact. There are steganographic methods that alter the media in such a way that in some cases the image can be processed to make the noise visible to the human eye. For example, least significant bit (LSB) embeddings cause significant randomizations of the pixel LSBs, randomizations that are obvious when the LSBs of the image are viewed alone. Without the presence of the steganographic data, the least significant bits of the pixels do have some correlation with the image content, whereas the presence of steganographic data will make the least significant bit plane look random. Highly textured images may also have a least significant bit plane exhibiting similar characteristics, though, so this method of steganalysis is not foolproof but is merely presented as an example.

Although data hidden within an image is often imperceptible to the human eye, the statistical nature of the image is disturbed. First attempts at steganalysis targeted the first order distributions of intensity or coefficients. As steganographic algorithms became smarter to circumvent the obvious statistical giveaways, the field of steganalysis has progressed to examining higher order statistical image characteristics. Image statistical analysis works as a steganographic method because natural image pixel interrelations are disturbed as a side effect of the embedding process. In the case of spatial domain embedding, the use of any of the bit planes for data hiding will most likely decrease the natural order and correlation that exists locally between neighboring pixels, neighboring bit planes, and neighboring color planes. Conversely, for robustness reasons, most frequency domain embedding techniques modify the low frequency components of the image, making the changes to the image globally rather than locally as with spatial techniques. Instinct would be to assert that frequency domain embedding techniques, while they disrupt some of the order within the media, tend to be statistically safer from spatial anomaly detection than spatial embedding techniques. This is demonstrated to be untrue, as the algorithm just as easily identifies transform domain modifications.

Many experiments have been conducted using image statistics and characteristics to discriminate between clean images and images with steganographic content. Among the many researchers are Avcibas, who has shown that both image quality metrics and binary similarity measures can be used for image steganalysis with fairly reliable results. Avcibas's techniques are based on the observation that natural images and images harboring steganographic data blur differently, thus treating the search for hidden data as a local, spatial process.

In (2) Avcibas uses binary similarity measures and ANOVA analysis to discriminate steganographic images from clean covers. Seven different similarity measures are selected based onto regression analysis and are used to compute statistics which together form the input vector for the neural net. Avcibas used a similar algorithm to examine the usefulness of image quality metrics for steganalysis, and obtaining a successful result.

Among the general algorithms, Farid developed a method for steganalysis based on multi-scale wavelet decomposition, using first and higher statistics to capture certain statistical regularities of natural images. Using a trained linear Fisher discriminant, Farid could detect data embedded with Jsteg-Jpeg, Outguess, and EZStego with greater than 95% accuracy.

1.3. Artificial neural networks for steganalysis

Artificial neural networks (ANNs) are recognized as powerful data analysis and modeling tools. They have been shown to capture and accurately represent both linear and non-linear relationships, and are an invaluable tool for approximating functions, clustering data, and recognizing patterns that are otherwise imperceptible. Neural networks can often be used in place of traditional statistical analysis methodologies such as time series models, regression analysis, ANOVA analysis, and traditional clustering techniques such as the K-means models. ANNs are very simple to apply to pattern recognition problems, requiring minimal knowledge about pattern recognition itself, and been used extensively in machine learning, knowledge discovery, and image analysis.

Artificial neural networks have been shown to produce highly accurate function approximation results, and can model non-linearities that are inherent in most pattern recognition problems, making the application of a neural net ideal in this setting. In (15), Shaohui demonstrated that a neural network can be trained to identify images harboring data embedded with any one of various watermarking and steganographic techniques, including Cox and Digimarc watermarks and Pretty Good Signature (PGS). The neural nets described by Shaohui were trained using the modeled distribution of the wavelet transform coefficients at each of the first three levels of the decomposition.

1.4. Local binary pattern operator

The side effects of the embedding process are manifested as small local variations in color or intensity within a small neighborhood in the color image. The local binary pattern operator was developed as a gray-scale invariant pattern measure that takes into account the amount of texture present in an image. It was first mentioned in (9) by Harwood and formally introduced in the form used in this paper for use in texture analysis in (14) by Ojala. The LBP is revered for both the computational simplicity and discrimination performance. The LBP operator also seems to correspond loosely to the pattern recognition methods in the human visual system (12).

The calculation for the local binary pattern value for pixel *p* uses the eight neighbors of *p*, together comprising a 3x3 square of pixels. The 0 to 255 intensity value for each pixel is obtained for each of the pixels in the square, and the outer pixel values are thresholded by the value of center pixel, *p*. An eight bit integer is composed from the outer thresholded values to formulate a LBP value for the center pixel p. A LBP value for each pixel in the image is placed into a 256-bin histogram and the histogram is stored along with a description of the texture it describes.

If the LBP were to be used for texture analysis, the histogram of the image in question would be compared to the stored histogram of known images and textures by using the log-likelihood ratio. Since our goal is steganographic analysis as opposite to texture analysis, instead of taking this step standard statistics describing the LBP histogram are calculated, including the standard deviation, variance, and mean. These statistical values are used in the input vectors to the artificial neural net.

Four different methods of calculating the LBP were used in the experiment to determine which was best for purposes of steganalysis. Two methods tried form a 9-bin histogram; two methods form a 256-bin histogram. The rationale was that possibly one method of computing LBP would be more advantageous in identifying hidden content. Of the 256 bin options, the first and coincidentally best methodology involved computing the LBP value by forming the thresholded bit values into an integer in a left to right fashion as depicted below.

The LBP value is the integer with the least significant bit defined by the value in the top left, and the most significant bit in the lower right

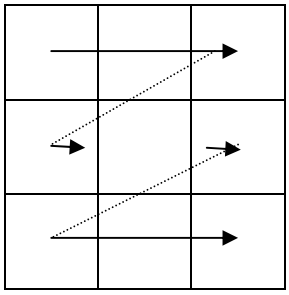

Figure 1: Left to Right Computation of the LBP

2. METHODOLOGY

2.1 Description of experiment

While multitudes of data can usually be hidden in an image, to help ensure imperceptibility it is recommended that the embedded bit rate be less than 0.02 bits per pixel (bpp) to minimize the probability of detection, especially for spatial embedding techniques. For this reason in the following experiment the steganographic algorithms where possible embedded less than this amount. For the test image sets, a randomly generated one hundred (100) byte sequence was used for the images to be modified, making a embedded bit rate of no more than 0.0082. For the training set, sixty (60) random bytes were inserted, or 0.0049 embedded bits per pixel.

While steganographic applications are specifically designed to hide data imperceptibly within media, the inclusion of steganographic noise into image data inevitably results in recognizable deviations from the natural pixel patterns present in clean images. If there is a good way to recognize steganographically-induced image texture patterns, the steganalysis problem could be solved using standard pattern recognition methodologies. Upon presentation of a feature set, a trained neural net could then discriminate between clean and steganographic images based on the texture pattern statistics. The local binary pattern operator was selected for this task because of the speed at which it could be computed and for its excellent texture analysis performance.

After computing LBP histograms for the image sets, a neural net was trained with statistics from 1000 clean color JPG images and 1000 images with one hundred (100) random bytes embedded using the Blindside steganographic program. The decision making capability of the neural net was tested by giving it input vectors from images modified using the F5, Digimarc, J1, PhaseMark, and JP Hide and Seek embedding algorithms.

Table 1: Steganalysis Results, Neural Net Trained with Blindside (100 bytes)

Image Set	Detection Rate
Clean Images	991/1000 (99.1%)
PhaseMark (strength 11)	886/1000 (88.6%)
JP Hide and Seek (100 bytes)	866/1000 (86.6%)
F5 (100 bytes)	865/1000 (86.5%)
J1	21/24 (87.5%)
Digimarc Watermark (Strength 1)	22/25 (88.0%)

The experiment was repeated with a training set consisting of statistics from clean images and images holding sixty bytes of data embedded with Blindside. The amount of hidden data in the training set is less than 0.005 bits per pixel. Because of the smaller amount of data hidden in the images, there should be fewer side effects of embedding, thus the input vectors for steganographic images should be closer to the input vectors of the clean images, and the performance of the neural net should decline.

Table 2: Steganalysis Results, Neural Net Trained with Blindside (60 bytes)

Image Set	Detection Rate
Clean Images	990/1000 (99.0%)
PhaseMark (strength 11)	829/1000 (85.2%)
JP Hide and Seek (100 bytes)	835/1000 (84.9%)
F5 (100 bytes)	831/1000 (83.1%)
J1	21/24 (87.5%)
Digimarc Watermark (Strength 1)	22/25 (88.0%)

Data uses as input features for the neural net include the delta between histogram bins and first order statistics derived from the histogram bins. Entries in the input vector were selected by examining the table below.

Table 3: Pearson Product Moment Correlation Coefficient (r)

Embedding Method	Color Plane	LBP Histogram Bin Statistics			
		Delta	**Mean**	**Variance**	**Entropy**
Blindside (60 bytes)	**R**	0.520	0.121	0.127	-0.154
	G	0.616	0.109	0.112	-0.212
	B	0.637	0.126	0.126	-0.243
Blindside (100 bytes)	**R**	0.499	0.042	0.040	-0.198
	G	0.598	0.043	0.042	-0.246
	B	0.618	0.050	0.056	-0.275
F5 (100 bytes)	**R**	-0.186	0.042	0.046	0.065
	G	-0.183	0.039	0.042	0.063
	B	-0.179	0.048	0.052	0.062
J1	**R**	0.194	-0.122	-0.130	-0.110
	G	0.188	-0.100	-0.108	-0.093
	B	0.204	-0.095	-0.103	-0.100
Digimark (strength 1)	**R**	0.024	0.067	0.065	0.072
	G	0.026	0.065	0.063	0.071
	B	0.024	0.061	0.057	0.069
JP Hide and Seek (100 bytes)	**R**	0.009	0.010	0.008	0.012
	G	0.009	0.011	0.008	0.013
	B	0.009	0.011	0.009	0.012
PhaseMark (strength 11)	**R**	-0.694	0.433	0.459	0.550
	G	-0.714	0.446	0.472	0.554
	B	-0.709	0.446	0.471	0.545

3. DISCUSSION

The presence of data within an image causes discrepancies in the image statistics, and the steganographic image can be discerned from a natural, clean image. Past work utilizing first order statistics for steganalysis was fruitful, and spawned a flurry of algorithms that embedded data within the image while preserving the first order statistics. Recent work concludes that natural images also have meaningful second order statistics. This has been taken into consideration in recent steganographic programs.

Besides using general image statistics, there may be texture characteristics that can be used to discern clean images from steganographic. This work shows that even images modified using frequency domain data hiding techniques exhibit texture characteristics that make the image discernable from a natural image. Additionally, passive steganalysis can be performed with reasonable accuracy on color images by using a texture-based feature set. We have shown that the statistics drawn from the LBP texture analysis process include effective statistical features that can be used for the purposes of steganalysis. One interesting result is the assertion that local, spatial texture analysis techniques can be used to correctly categorize images harboring transform domain embedded data.

Neural networks tend to be useful when it comes to identifying and categorizing patterns. Interestingly, the neural net was trained to identify images modified using Blindside, and the neural net successfully identified most images modified with other algorithms included. It should be mentioned that the experiment was repeated using other image sets as the neural net training set, only with much less successful results. Additionally, images carrying 0.005 bits of hidden data per pixel have texture characteristics distinct enough from natural images that the texture features can still be used to train a neural net to recognize other steganographic images.

Extensions of the research presented would be furthering the examination of texture analysis for the purposes of steganalysis to find identifiable feature sets to identify different embedding techniques. If such a feature set could be found it would be interesting to see if a self organizing map could be created, grouping together images altered with similar embedding methodologies, i.e. least significant bit, discrete cosine transform techniques, and others.

REFERENCES

1. I. Avcibas. *Image Quality Statistics and their Use in Steganalysis and Compression*. Ph.D. Thesis, 2001.
2. I. Avcibas, N. Memon, and B. Sankur. *Image Steganalysis with Binary Similarity Measures*, ICIP, Rochester, Vol. 3, 645-8, September, 2002.
3. I. Avcibas. *Steganalysis Based on Image Quality Metrics*. IEEE Fourth Workshop of Multimedia Signal Processing, Pp. 517-522, 2001.
4. I. Avcibas, B. Sankur, N. Memon, *Steganalysis of Watermarking and Steganogarphic Techniques Using Image Quality Metrics*. IEEE Transactions on Image Processing 12(2), 21-229, 2003.
5. G. Berg, I. Davidson, M. Duan and G. Paul. *Searching For Hidden Messages: Automatic Detection of Steganography*. Innovations of Artificial Intelligence , 2003.
6. H. Farid. *Detecting Hidden Messages Using Higher-Order Statistical Models*. IEEE International Conference on Image Processing, 2002.
7. H. Farid. *Detecting Steganographic Messages in Digital Images*. Technical Report, TR2001-412, Dartmouth College, 2001.
8. J. Fridrich, and M. Goljan. *Practical Steganalysis of Digital Images – State of the Art*. Security and Watermarking of Multimedia Contents, , Vol. SPIE-4675, Pp. 1-13, 2002.
9. D. Harwood, T. Ojala, M. Pietikäinen, S. Kelman and L. Davis. *Texture Classification by Center-Symmetric Autocorrelation, using Kullback Discrimination of Distributions*. Pattern Recognition Letters 16:1-10, 1993.
10. J. Li, and J. Wang. *Automatic Linguistic Indexing of Pictures by a Statistical Modeling Approach*. IEEE Transactions on Pattern Analysis and Machine Intelligence, volume 25, number 9, Pp. 1075-88, 2003.

11. S. Lyu and H. Farid. *Steganalysis Using Color Wavelet Statistics and One-Class Support Vector Machines*. SPIE Symposium on Electronic Imaging, 2004.
12. T. Maenpaa. *The Local Binary Pattern Approach to Texture Analysis – Extensions and Applications*. Ph.D. Dissertation, University of Oulu, 2003.
13. R. Newman, I. Moskowitz, L. Chang, M. Brahmadesam. *A Steganographic Embedding Undetectable by JPEG Compatibility Steganalysis*. Lecture Notes in Computer Science, Papers from the Fifth International Workshop on Information Hiding. Pp. 258-77, 2002.
14. T. Ojala, M. Pietikäinen, and K. Harwood. *A Comparative Study of Texture Measures with Classification based on Feature Distributions*. Pattern Recognition 29: 51-59, 1996.
15. L. Shaohui, Y. Hongxun, G. Wen. *Neural Network Based Steganalysis in Still Images*. IEEE International Conference on Multimedia and Expo, July 6-9, 2003.
16. L. Shaohui, Y. Hongxum, G. Wen. *Steganalysis Based on Texture Analysis and Neural Network*. Proceedings of WCICA2004, HangZhou, China, 2004.
17. J. Wang, L. Jia, and G. Wiederhold. *SIMPLIcity: Sematics-Sensitive Integrated Matching for Picture LIbraries*. IEEE Transactions on Pattern Analysis and Machine Intelligence, volume 23, number 9, Pp 947-63, 2001.