Forensic Image Recognition using a Novel Image Fingerprinting and Hashing Technique

R D Neal, R J Shaw and A S Atkins

Faculty of Computing, Engineering and Technology, Staffordshire University, Stafford ST18 0AD, UK

This paper identifies a problem within forensics computing entailing the recognition of known suspect images which may have been modified. Current hashing techniques mean that if an image is changed by its file type, it cannot be detected from the known hash of the original image file. The technique outlined in this paper solves this problem by utilising techniques used in audio recognition algorithms to identify greyscale and colour images. This technique is also able to output as a percentage, the match between the original image and any modification of the image. The technique is able to identify suspect files that may require further investigation and so increases the efficiency of the pre-investigation stage. It can be used against a wide range of image modification techniques that can be used to avoid detection. The paper outlines various recognition methods that can be used to detect suspect files located within electronic media.

Index Terms **– Digital Forensics, Fourier Transforms, Image Fingerprinting, Image Content Analysis;**

I. Introduction

A new technique has been proposed [1] and this paper uses this technique to develop a new and robust method for locating suspect images during a digital forensic investigation of digital images in colour.

The new image analysis method described in this paper has been tested by comparing original images with modified versions of the same image. Modifications of interest in this paper include inversion, pixel modification, noise, blurring, edge extraction, colour modification, brightness adjustment, greyscale and non-matching images.

The initial hypothesis is derived from that of a novel hashing algorithm for images. The hypothesis was formed because current forensic investigations use a "hash matching" technique to determine the presence of any suspect files. This technique assumes that there has been no modification to the suspect file, from the original and so hashing techniques do not account for change to the file. The hash of a file is usually taken during a forensic investigation to ensure the integrity of the file and that evidence had not been compromised. This means that if a file is modified in any way, then the hash will be changed. Due to the implementation of commonly used hashing algorithms, such as MD5 and SHA-1, the hashes will be unrelated. This means it would be impossible to detect what has been modified without the original file for comparison [1].

Two modern algorithms approved by the National Institute of Standards and Technology (NIST) are the ***ssdeep*** and ***sdhash***. These two hashes have been evaluated in order to review the problem outlined. Both of these hash algorithms are Context Triggered Piecewise Hashes (CTPH), which means that they are developed for detection of homologous files by producing hashes for pieces of the files rather than the entire file collectively. This means that multiple hash values must be stored for one file, increasing the amount of data which is needed to be stored. Using these techniques it is possible to detect modifications within an image, but not the specific area or pattern of change within that image. An example of this can be seen in Figure 1. It can be seen that one of the hashes is different in the modified file, indicating a change within that section of the file [2].

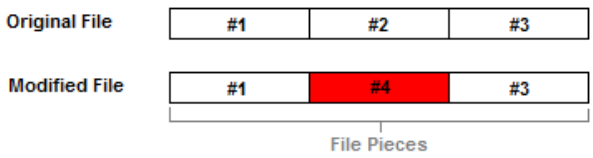

Figure 1 - CTPH change detection

The ***ssdeep*** algorithm produces a fixed length hash. This means it can always be determined the size of the hash from the size of the file. The ***sdhash*** algorithm produces a static length hash value, and so the size of the hash cannot be calculated. This algorithm has a different approach to ***ssdeep***, as it utilises the ***Bloom Filter*** algorithm in order to determine if there is a modification. Both of these algorithms can present false positives (duplicate values for unrelated or modified files) [2].

The experiments detailed in this paper address these issues by being able to determine more specifically what data has been modified in an image by creating a small audio fingerprint of the file. This technique is a powerful and robust image analysis technique which would provide both a hashing and content analysis algorithm. As opposed to ssdeep or sdhash, the algorithm would work for multiple image file types and so is acceptable as a hashing algorithm for any image files found on a file system. The purpose of the proposed algorithm is to help identify identical or modified image content, as opposed to a data comparison. For the experimental stages of this paper, two identical images with different file types (jpg and png) were hashed using ssdeep.

		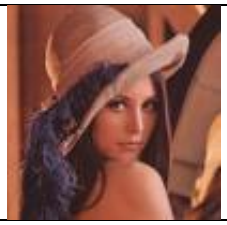
Type	PNG	JPG
Hash Block	XIHMDC/546wkT4Kx5C+/IukNjGUKFuGljwDpmP85LfF27WyrkATLOKaeq2gsdEMD	Zx8F7rrrbqNUDiJqZo5ltIPldcOOwwVP3z03e7uJZBa3F9Jj9iobuu+t9pKEuZ
Hash Block	XIskTbx33UK84jwDpmP6LfOL6F0dEMt9	3arqiiJqdcOgPz0uYBa19Gampg

Table 1 - ssdeep Image Comparison

As can be seen from Table 1, when comparing the hash values obtained, it is clear that there is no match between the two files, although visually being identical. The actual ***ssdeep*** program cannot make a comparison between the two files because they are not the same file type. The proposed solution in this paper will be able to perform such a comparison irrespective of file type

II. EXPERIMENTAL

The complexity of transforming an image to sound and then performing pattern recognition on the generated audio, involves performing a Fourier transform in order to fingerprint the data. By performing a Two-Dimensional (2D) Fourier Transform using an image a 2D matrix of transform information can be generated. Audio fingerprinting utilises a One-Dimensional (1D) Transform to fingerprint data [3]. From this theory, a current technique can be modified to also work for 2D transforms. As audio fingerprinting techniques are an already proven and robust method of pattern recognition, it is theorised that a modified technique will provide the benefits to produce improved results in this paper. An analysis of both techniques was reviewed and the root-mean-square (RMS) error results compared. The current process can be seen in Figure 2.

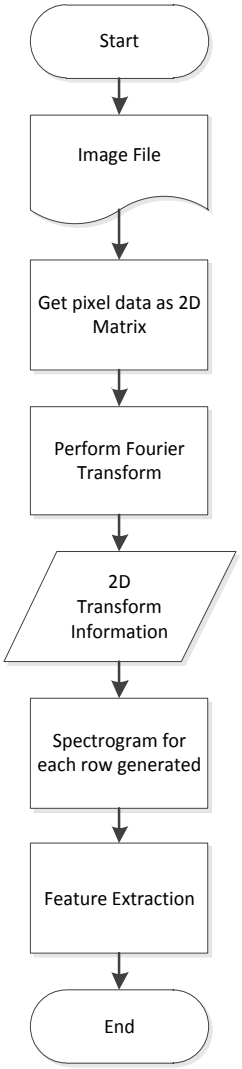

Figure 2 - Image Feature Extraction

The algorithm which best suits the needs of the technique described in this paper was required to be determined. In order to do this, a comparison of the three audio fingerprinting algorithms was performed. The experiment compared False Negative (FN) (not recognising a song) and False Positives (FP) (identifying the wrong song), as well as its ability to recognise audio that had been altered (distorted), measured as a Robust Recognition (RR) rate calculated by the mean percent of successful recognitions. It must be noted that both FN and FP rates are both classified as test fails for calculating the RR rate. Successful Recognition (SR) rate is a measure of the recognition of un-distorted audio. This rate is calculated as:

$$SR = 100 - (FN + FP)$$

Existing mobile applications were used for the experiment. Ten songs were attempted ten times with each implementation. The results are shown in

Table 2. The algorithms compared were Philips Robust Hash (PRH), Shazam (Avery-Wang) and Query by Humming (QbH) [3].

	FN rate	FP rate	SR rate	RR rate
PRH	4%	6%	90%	89%
Shazam	3%	5%	92%	92%
QbH	0%	8%	92%	68%

Table 2 - A comparison of audio recognition techniques

All the results of the experiment showed relatively low FN and FP error rates, proving that these algorithms are accurate audio recognition techniques. However, a noticeable difference was found for the RR for distorted audio between the different algorithms.

For the PRH algorithm, the results show a 90% SR rate and a 1% difference between the SR and RR rates. This shows that some functionality is lost when the audio is distorted. With respect to the Shazam and QbH algorithms, the FP rate is average, meaning it is not the weakest choice to be considered. The FP rate is the most important measure, since within forensic analysis it is inappropriate to wrongly identify an innocent image as malicious. From this, an algorithm with a low FP rate should be chosen. It is equally as important that the FN rate is low because a malicious image is desired to be detected, which may not be feasible with an algorithm with a high FN rate. The FN rate for PRH is 4%, which is high with respect to the Shazam and QbH algorithms. This rate is still a small statistically, but shows the likelihood of FN results occurring is greater using the algorithm.

The results of the QbH algorithm show that there is a drastic difference between the SR and RR rate of 24%. This is an exceptionally large difference in comparison to the Shazam and QbH algorithms. Consequently, the QbH algorithm may not be suitable for detecting certain methods of image alteration and so may not be suitable as a solution. The FP rate was also the highest of the three algorithms, showing that the algorithm may again not be suitable. The FN rate of the algorithm (for none distorted audio) was 0% which is the lowest of the three algorithms, showing that it is very good at matching its fingerprints to that of one of its database entries.

The results of the Avery-Wang (Shazam) algorithm show that there is a 92% SR rate for song identification. The RR rate is also 92%, indicating there is no loss to functionality, when compared to the SR rate, by applying distortion to the audio. This shows that it is the most robust algorithm against distortion out of the algorithms tested. The FP rate is 5%, which is low with respect to the Shazam and QbH algorithms. Consequently, it may be prudent to choose the Shazam algorithm to provide a solution. This is supported by the FN rate which is 3%, which is the second highest of the three algorithms compared. Ideally the SR rate should be 100%; consequently the Shazam algorithm produced the lowest combined error rates (FN + FP) with the best RR rate, showing it is the strongest algorithm tested. From this analysis, the Shazam algorithm was chosen for the experimentation.

A spectrogram's purpose within the Shazam algorithm is the process of extracting points of minutiae following a Hamming Window function. The theory behind an audio spectrogram shows that a 1D matrix (audio data) results in a 2D matrix (an image of a spectrogram). This paper focuses on the content analysis of imagery and so it must be applied to the starting input of a 2D matrix. It is already common practise for 2D transforms to be applied to imagery. Spectrograms utilise the Discrete Fourier Transform (DFT) [4]. The theory was applied to process each row individually in order to generate a list of spectrograms, which would result with a Three-Dimensional (3D) matrix. This theory can be seen in Figure 3.

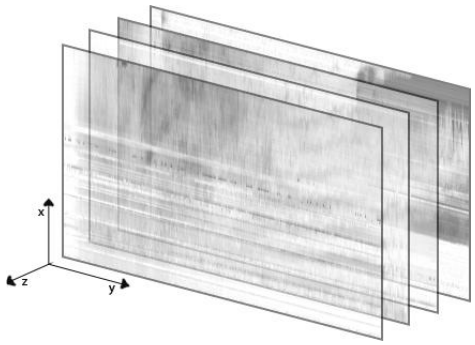

Figure 3 - 3D Spectrogram

The key points for fingerprinting were extracted by calculating the highest frequencies in the spectrogram [5] and setting those pixels to black. All other pixels were set to white. The difference rate can then be calculated by comparing spectrograms pixel-by-pixel to match for the black points. If a key point on one fingerprint does not match on the other fingerprint, then the result is a 0% match, and if it does match then it is registered as 100% match. All white pixel matches are ignored. The match percentage can then be calculated by taking the amount of fails, dividing it by the total matches and fails, and multiplying by 100, and subtracting this value from 100.

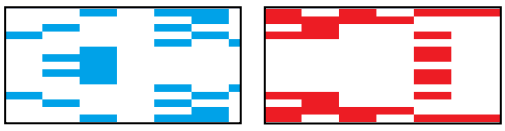

Figure 4 - Feature extraction of a row of two images

Figure 4 shows the extracted features of two images using the technique described. These key points are matched against each other in order to calculate the amount of matches to failed matches. An example can be seen in Figure 5 where the matches are represented as green, and all non-matches are blue or red; white pixels are ignored. The match rate for these example rows is 23%. An average percentage for all the rows can be calculated in order to deduce the overall match for the image.

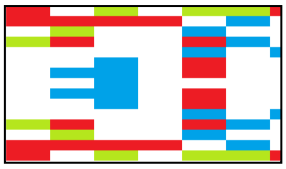

Figure 5 - Overlay of features from Figure 4. Matches represented as green

From the acquired match rates for different image modification techniques, a threshold will be deduced in order to state what match percentage can indicate an accurate match (e.g. < 95% match is not classed as an accurate match). This will be done by determining first that the algorithm successfully matches an identical image, and then discovering the match rate for non-matching images. This will determine an initial threshold which can then be used against modified images in order to determine which modifications the technique is robust to.

III. Results

The technique described in the paper was originally designed to be able to detect images that could be verified by human perception but not by simple image analysis techniques. A change to the colour of an image (e.g. colour to greyscale) would be easily analysed by an examiner, but not by a simple pixel-matching algorithm. The technique in this paper would be able to be automated to search a suspects acquired files (which may be several GB of data) and be able to detect suspicious images known in the investigators image fingerprint database. In order to establish how robust the technique is, the technique will be applied to common image modification methods, and the results with give an indication to how successfully this method will recognise the image.

In order to determine accurate thresholds for matches, the technique was applied to 10 images. For the applicable modification techniques, several different levels were applied to each image and compared to the original. The lowest match results and highest match results were noted and can be seen in

Table 3.

Modification	Highest Match Rate	Lowest Match Rate
Increased Brightness	98%	88%
Decreased Brightness	97%	93%
Inversion	100%	100%
Blurring	98%	97.5%
Edge Extraction	73%	51%
Greyscale	94%	90%
Noise	91%	65%
Colour Adjustment	95%	62%
Pixilation	98%	94%
JPEG Compression	97%	95%
Rotation	94%	70%

Table 3 - Thresholds for Image Modifications

From the results of non-matching images, it can be deduced that the threshold for matching an image is 80%, due to the results for the highest match for any non-matching image. If this is cross-referenced with the lowest match for any of the other modification technique, it is possible to determine how robust the experiment is to modification techniques. This can be seen more clearly in Figure 6. The proposed method cannot be relied upon when the matching rate for the modification technique used is below 80 % (Threshold). The results show that Edge Extraction, Noise, Colour Adjustment and Rotation were below the match rate threshold and so are considered not to be recognisable using this technique.

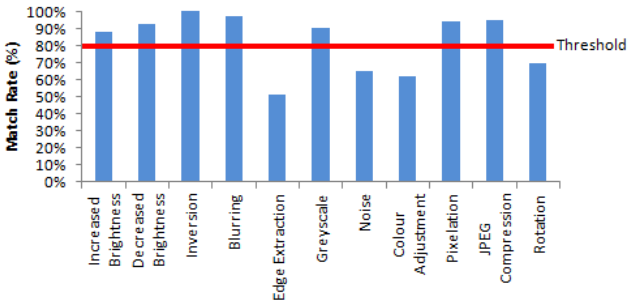

Figure 6 - Experiment Threshold Comparison

Figure 7 shows the results from a set of three rotations of 20^{o}, 45^{o} and 90^{o}. Each rotated image was matched against the original un-rotated template. The results indicate that for an image rotation of $<20^{o}$, the image is recognisable and a match is found.

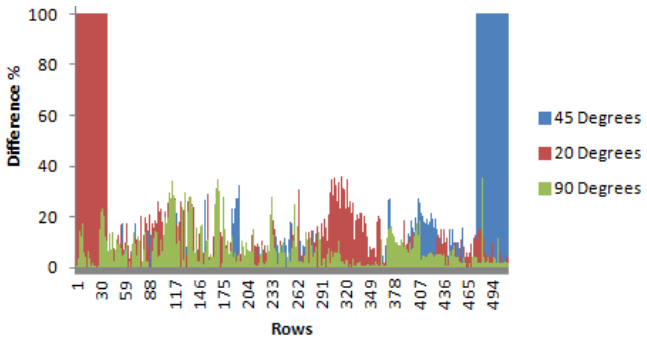

Figure 7 – Experimental Rotation Comparison

This low accuracy of recognition of rotated images can be overcome. It can be achieved by generating a set of images, each of an incremental rotation of 20^{o}. This gives a total of 18 rotated images that can be stored as templates for matching purposes of suspected files.

IV. Conclusion

The research addressed hashing techniques and how they are not robust to visual imagery. A solution was proposed to utilise audio recognition techniques in order to fingerprint the images. These image fingerprints would serve as a hash to identify images as having the same or similar content, and

hence serve a better purpose than that of piecewise hashing. Audio recognition techniques were analysed, and an experiment was generated to determine which method would be most suitable for the purpose of this paper. The results concluded that the Shazam algorithm would be most effective.

Utilising the techniques used in the Shazam algorithm, an experiment was generated which resulted in generating fingerprint hashes for images input into the experiment program. These fingerprints were then compared. Multiple scenarios were given for comparison involving different image modification techniques in order to evaluate how robust the fingerprinting technique served for hashing. The results showed that it is a lot more effective than piecewise hashing, due to the fact that it could be calculated as a percent how much of a difference the modified image was compared to the original.

The results of the experiment showed that the technique was robust to matching non-modified images with little to no difference, and was also capable of matching inverted images with no difference at all. It also showed high match rates for darkened/lightened images, JPEG compressed images, blurred, pixelated and greyscale images. This means that the technique could be utilised in forensic analysis for identifying these types of modified malicious image files.

The experiment results found the technique to be less robust to lightened (above 60% brightness), colour adjusted, noisy, high passed and rotated images. This means that the technique is not fully functional for all image modification techniques. This may be because image modification techniques utilise different image processing techniques in order to generate the modification.

It would be desirable to use the same technique using a different signal transform, such as the Discrete Cosine Transform or Discrete Wavelet Transform. Using these techniques, it may present more accurate results.

A method for increasing the size of the spectrograms for fingerprinting could be used in order to increase the key points for fingerprinting, and as a result retrieve more accurate results. It is proposed that all the image data be placed together in order to generate one large spectrogram as opposed to multiple small spectrograms.

REFERENCES

[1] R. J. Shaw and A. S. Atkins, "A Novel Way of Identifying Suspect Picture Files," International Conference in Cybersecurity, Cyberwarfare & Digital Forensics (CyberSec12), UCTI, Malaysia, June 2012 (in press)

[2] V. Roussev, "An evaluation of forensic similarity hashes," *Digital Investigation*, vol. 8, p. S34-S41, Aug. 2011.

[3] H. A. V. Nieuwenhuizen, W. C. Venter, and L. M. J. Grobler, "Comparison of Algorithms for Audio Fingerprinting," pp. 5-6, 2011.

[4] Q. Wang, G. Liu, Z. Guo, J. Guo, and X. Chen, "Structural Fingerprint Based Hierarchical Filtering in Song Identification Pattern Recognition and Intelligent System Laboratory ," *Beijing University of Posts and Telecommunications*, pp. 1-4, 2011.

[5] B. Zhu, W. Li, Z. Wang, and X. Xue, "A novel audio fingerprinting method robust to time scale modification and pitch shifting," in *Proceedings of the international conference on Multimedia - MM '10*, 2010, p. 987.