REASSEMBLY AND CLUSTERING BIFRAGMENTED INTERTWINED JPEG IMAGES USING GENETIC ALGORITHM AND EXTREME LEARNING MACHINE

RABEI RAAD ALI

A thesis submitted in
fulfillment of the requirement for the award of the
Degree of Doctor of Philosophy

Faculty of Computer Science and Information Technology
Universiti Tun Hussein Onn Malaysia

JULY 2019

This thesis is dedicated to:

The sake of Allah, my Creator.

My great teacher and messenger, Mohammed (May Allah bless and grant him), who taught us the purpose of life;

My great parents, who lead me through the valley of darkness with the light of hope and support;

My beloved brothers and sisters;

To all my family, the symbol of love and giving;
My friends who encourage and support me;

All the people in my life who touch my heart;

I dedicate this research.

ACKNOWLEDGEMENT

Firstly, all praise and thanks to Allah for His Divine Guidance and help in completing this research project.

My Salaam and Gratitude to the Beloved Prophet Mohammed (Peace and Blessings of Allah Be Upon Him) who was sent by Allah to be a great teacher of human kind.

I would like to thank my Supervisor, Dr. Kamaruddin Malik Bin Mohamad for his encouragement, ideas, and support throughout the research project. I appreciate and cherish the moments I have spent with him since the start of this research. We have had productive discussions and constructive arguments over the theories and concepts proposed in this thesis.

I thank all those who have supported me in this endeavor: my friends, the department, colleagues, fellow students and members of Faculty of Computer Science and Information Technology (FSKTM).

Finally, I would like to thank the Office for Research, Innovation, Commercialization and Consultancy (ORICC), Universiti Tun Hussein Onn Malaysia (UTHM) for providing the funds to conduct my Doctoral research under Project Vot No. U495.

ABSTRACT

File carving tools are essential element of digital forensic investigation for recovering evidence data from computer disk drives. Today, JPEG image files are popular file formats that have less structured contents which make its carving possible in the absence of any file system metadata. However, completely recovering intertwined Bifragmented JPEG images into their original form without missing any parts or data of the image is a challenging due to the intertwined case might occur with non-JPEG images such as PDF, Text, Microsoft Office or random data. In this research, a new carving framework is presented in order to address the fragmentation issues that often occur in JPEG images which is called RX_myKarve. The RX_myKarve is an extended framework from X_myKarve, which consists of the following key components: (i) an Extreme Learning Machine (ELM) neural network for clusters classification using three existing content-based features extraction (Entropy, Byte Frequency Distribution (BFD) and Rate of Change (RoC)) to improve the identification of JPEG images content and support the reassembling process; (ii) a genetic algorithm with Coherence Euclidean Distance (CED) matric and cost function to reconstruct a JPEG image from a set of deformed and fragmented clusters in the scan area. The RX_myKarve is a framework that contains both structure-based carving and content-based carving approaches. The RX_myKarve is implemented as an Automatic JPEG Carver (AJC) tool in order to test and compare its performance with the state-of-the art carvers such as RevIt, myKarve and X_myKarve. It is applied to three datasets namely DFRWS (2006 and 2007) forensic challenges datasets and a new dataset to test and evaluate the AJC tool. These datasets have complex challenges that simulate particular fragmentation cases addressed in this research. The final results show that the AJC with the aid of the RX_myKarve framework outperform the X_myKarve, myKarve and RevIt. The RX_myKarve is able to completely carve 23.8% images more than X_myKarve, 45.4% images more than myKarve and 67% images more than RevIt in which AJC tool using RX_myKarve completely solves the research problem.

ABSTRAK

Peralatan ukiran fail (*file carving*) adalah merupakan elemen yang penting dalam penyiasatan forensik digital bagi memulihkan data bukti storan cakera komputer. Pada masa ini, fail JPEG adalah merupakan format fail yang popular yang mempunyai kandungan yang kurang berstruktur yang membolehkan process ukiran fail boleh dilakukan tanpa adanya sebarang metadata sistem fail. Walau bagaimanapun, pemulihan fail JPEG yang telah dipecahkan kepada beberapa fail atau serpihan (*fragmented*) merupakan suatu cabaran dan tidak mudah untuk diatasi memandangkan kerumitan proses mendapatkan semula serpihan imej JPEG terutamanya intertwined JPEG. Dalam penyelidikan ini, sebuah rangka kerja baharu bagi *carving* dibentangkan bagi menangani isu serpihan (fragmentation) yang sering berlaku dalam fail imej JPEG yang dipanggil RX_myKarve. RX_myKarve adalah rangka kerja tambahan terhadap X_myKarve, yang terdiri daripada komponen utama berikut: (i) sebuah rangkaian neural mesin pembelajaran Ekstrim (*Extreme Learning Machine* atau ELM) yang mengklasifikasi kelompok untuk menambahbaik proses mengenal pasti kandungan imej JPEG dan menyokong proses pencantuman semula; (ii) Algoritma Genetik untuk membina semula imej JPEG daripada set kluster yang rosak dan berpecahan dalam *scan area.* RX_myKarve adalah sebuah rangka kerja yang komprehensif yang mengandungi kedua-dua ukiran berasaskan struktur dan pendekatan ukiran berasaskan kandungan. RX_myKarve dibangunkan sebagai alat Automatic JPEG Carver (AJC) bagi menguji dan membandingkan prestasinya dengan *carving tool* seperti RevIt, myKarve dan X_myKarve. Ianya telah diuji pada dataset DFRWS (2006 dan 2007) dan satu dataset baru. Dataset-dataset ini mempunyai cabaran kompleks yang mensimulasikan kes-kes serpihan tertentu seperti yang dinyatakan dalam penyelidikan ini. Hasil akhir menunjukkan bahawa AJC dengan bantuan rangka kerja RX_myKarve mengatasi X_myKarve, myKarve dan RevIt. RX_myKarve dapat mengukir gambar 23.8% lebih imej berbanding X_myKarve, 45.4% lebih imej berbanding myKarve dan

67% lebih imej berbanding RevIt, di mana AJC dengan menggunakan RX_myKarve dapat menyelesaikan sepenuhnya masalah imej *Bifragmented JPEG*.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

LIST OF ALGORITHMS

LIST OF SYMBOLS AND ABBREVIATIONS

ADB	-	Address Database
AJC	-	Automated JPEG Carver
APF	-	Adroit Photo Forensics
APP	-	Application segment
AWQ	-	Automated Work Queue
BFA	-	Byte Frequency Analysis
BFD	-	Byte Frequency Distribution
BGC	-	Bifragment Gap Carving
BMP	-	Bitmap
CCITT	-	Consultative Committee on International
CED	-	Coherence of Euclidean Distance
CERT	-	Computer Emergency Response Team
COM	-	Comment
DC	-	Direct Current
DCT	-	Discrete Cosine Transform
DFRWS	-	Digital Forensics Research Conference
DHP	-	Define Hierarchical Progression
DHT	-	Define Huffman Table
DLN	-	Define Number of Lines
DOC	-	Microsoft Words
DPCM	-	Differential Pulse Code Modulation
DQT	-	Define Quantization Table
DRI	-	Define Restart Interval
ELM	-	Extreme Learning Machine
EOF	-	End of File
Exif	-	Exchangeable image file format
EXT	-	Extended file system

FAT	-	File Allocation Table
FDCT	-	Discrete Cosine Transform
FHT	-	File Header / Trailer
GIF	-	Graphics Interchange Format
IDCT	-	Inverse Discrete Cosine Transform
IEC	-	International Electrotechnical Commission
ISO	-	International Organization of Standard
ITU	-	International Telecommunication Union
JEIDA	-	Japan Electronics Industry Development
JFIF	-	The JPEG File Interchange Format
JPEG	-	Joint Photographic Experts Group
MCU	-	Minimum Coding Unit
MFC	-	Multimedia File Carving
MLP	-	Multi-Layer Perceptron
PCA	-	Principle Component Analysis
PDF	-	Portable Document Format
PUP	-	Parallel Unique Path
RGB	-	Red Green Blue
RoC	-	Rate of Change
RST	-	Restart interval Termination
SOF	-	Start of Frame
SOI	-	Start of Image
SOS	-	Start of Scan
SPP	-	Sequential Pixel Prediction
SVM	-	Support Vector Machine
UHP	-	Unique Hex Patterns
VJM	-	Validated JPEG Markers

LIST OF APPENDICES

LIST OF PUBLICATIONS

(i) **Ali, R. R.**, Mohamad, K., Jamel, S., & Khalid, S. K. A. (2018). A Review of Digital Forensics Methods for JPEG File Carving. Journal of Theoretical and Applied Information Technology, 96(17).

(ii) **Ali, R. R.**, Mohamad, K. M., Jamel, S., & Khalid, S. K. A. (2018). Classification of JPEG Files by Using Extreme Learning Machine. In International Conference on Soft Computing and Data Mining (pp. 33-42). Springer, Cham.

(iii) **Ali, R. R.**, Mohamad, K., Jamel, S., & Khalid, S. K. A. (2018). Extreme Learning Machine Classification of File Clusters for Evaluating Content-based Feature Vectors. International Journal of Engineering and Technology, 7 (4.36).

(iv) **Ali, R. R.**, & Mohamad, K. M. (2019). RX_myKarve Carving Framework for Reassembling Complex Fragmentations of JPEG Images. Journal of King Saud University-Computer and Information Sciences.

CHAPTER 1

INTRODUCTION

1.1 Overview

Studies show that the number of people that are using digital devices such as smartphones and computers from 2000 to 2017 have been increased up to 49.6%. This tremendous growth results from the technology advancements of high internet speed and large storages capacity (Kenney & Gortmaker, 2017). Subsequently, the multimedia files of digital images and documents have become the current trends in retaining important information or memories of digital devices (Pahade *et al.*, 2015). In some cases, the files are exposed to deformation or damage due to many reasons including device failure, deliberate destruction or human errors. Moreover, the digital forensic investigation encounters purposely deleted data of criminal actions such as terrorism, stolen goods, child pornography and theft (De Bock & De Smet, 2016).

The recovery of deformed Joint Photographic Experts Group (JPEG) images (damaged, fragmented or deleted) is a very important issue to the related users (Pahade *et at.*, 2015). Recovery of data files including JPEG, Text, Microsoft Office and etcetera when their system information missing is a challenging research issue (Uzun & Sencar, 2015; Tang *et al.*, 2016). The recovery process entails methods that analyze the structure and contents of each individual file blocks / clusters. Conventional data recovery methods use file system information (metadata) to recover such files. In the case of the absence of the file system metadata or the damage of file system itself, conventional methods of data recovery are unable to recover these files (Abdullah *et al*, 2016). As an alternative, few carving techniques are proposed to deal with the case of the absence of the file system metadata or the damage of the file system.

There are three categories of files' carving approaches: signature-based, structure-based and content-based. Each category has a number of advantages and disadvantages. Therefore, none of the categories are perfect and can provide comprehensive solutions (Qiu *et al.,* 2014).

A signature-based carving is a straightforward approach that has been successfully proven to carve contiguous files. This approach works on the header - footer data of the image i.e., header and footer (Nadeem Ashraf, 2013). However, in many cases, the signature block / cluster is damaged or disconnected due to storage damage or system fragmentation process (Uzun & Sencar, 2015; Hilgert *et al.,* 2019). Identifying the files type is an essential step to recover the files with missing or damaged file system information (Amirani *et al.,* 2013). As a result, it lacks handling fragmented data in both consecutive, contiguous, non-consecutive, and non-contiguous order cases as explained in Section 2.5 (Nadeem Ashraf, 2013).

A structure-based carving approach has been used to carve the fragmented file, by identifying and deleting the fragmented portion of the file in the scan area (Metx & Mora, 2006). This approach does not cover fragmented data files that have some non-contiguous and/or non-linear order cases (Kloet, 2007).

A content-based carving approach attempts to handle some cases of fragmented data files. This approach recovers files by analyzing the contents of the scan area. The carving process includes identification, indexing, matching, and reassembling the fragmented portions of the files. The content-based approach of file carving is still in its preliminary stages and not fully explored (De Bock & De Smet, 2016). There are only a limited number of carving methods that reconstruct files based on the content of the scan area of images (Li *et al.,* 2011). For example, Karresand & Shahmehri (2008), Abdullah *et al.* (2016) point out the importance of focusing on the fragmentation problem of the scan area in which the fragmentation point is located after Start of Scan (SOS) marker. Because, any non-JPEG clusters after the Start of Scan (SOS) can cause image distortion or corruption. Therefore, this thesis focusses on recovering of JPEG images with the absence of the file system metadata. The study adopts file carving as an advanced approach that recovers files in the absence of file system metadata. Through file carving, file contents of the scan area can be recovered as long as they are not corrupted or overwritten. The JPEG images are very suitable for carving due to their inherent file structure (Lewis, 2012). The aim of this research is to provide a reassembling technique to solve fragmentation scenarios in the scan

area of a JPEG image with/out Restart marker (RSTm) and a JPEG image with/out thumbnail(s) in which the JPEG image is split into two fragments (Bifragment) as maximum and/or a JPEG image intertwined with other non-JPEG images (i.e., the intertwined Bifragmented JPEG images) in the scan area. The following section presents the research motivation. The rest of the chapter presents a research problem, objectives, scope and research organization.

1.2 Research Motivation

Nowadays, digital forensics aims to provide an assistant for making the decision about a crime by analyzing data evidence and then looking at a file content which usually involves image files. There are many image formats such as Graphics Interchange Format (GIF), Bitmap Image File (BMP) and A Joint Photographic Experts Group (JPEG). A Joint Photographic Experts Group (JPEG) files are very popular because they have the features of easily compressed data, speed up transferring processes and efficient for internet in terms of speed and bandwidth (Li *et al.,* 2011). It is also one of the image file formats that has less structured contents which makes its recovery possible in the absence of file system metadata. Additionally, a lot of files on computers or smartphones are valuable such as documents and memorial photos. The files are exposed to unintentionally "permanently" deletion by end-users or corruption, by malware or hardware failure for instance (Hand *et al.,* 2012).

The modern operating systems store files contiguous, but a missing file can easily become fragmented due to the operating system fragmentation process. Kloet (2007) defines a fragmented file as a file that has been split into two or multiple parts where all parts are stored in different locations (non-contiguous) on a hard disk. Although, the percentage of fragmented files is relatively small (10%) these are usually the files that are of interest for forensic purposes (Garfinkel, 2007). Reassembling of fragmented file in files recovery is a hard task and not all recovery methods can handle. The fragmented files entail advanced recovery methods to identify, add, delete, match, and link files clusters (Xu & Dong, 2009; Durmuset *et al.,* 2019).

Karresand & Shahmehri (2008) introduced a technique to reassemble fragmented baseline sequential JPEG images using RST marker. The reassembly is performed by measuring the validity of a joint between two fragments. Qiu *et al*.

(2014) propose a new Multimedia File Carving (MFC) method using Parallel Unique Path (PUP) reassembling technique. The aim of the MFC method is handling high entropy file fragments with high recovery accuracy. File carving, in general, has three main procedures which are the identification of a file information, verification for the collected information of the file, and reassembly of the file (Tang *et al*., 2016). De Bock & De Smet (2016) define reassembling as a process of detecting a fragmentation point of a fragmented file and then starting point of the following fragment.

In terms of a digital forensic tool of the reconstruction of fragmented objects problem, which we call reassembling fragmented files has received little attention. Therefore, there are limited number of reassembling methods that reconstruct files from a collection of randomly mixed fragments. Most digital forensic tools concentrate on carving contiguous fragments of a fragmented data file with many false positive of their results (Abdullah *et al.,* 2016). For instance, Mohamad *et al.* (2010) propose myKarve as a technique to carve fragmented JPEG images caused by other file formats such as Word, PDF, and Excel which are called Garbage. Carrier (2005) propose a Digital Forensics Research Workshop DFRWS (2006) dataset that includes files with different carving challenges including contiguous and fragmented files. It is used by research for reassembly of fragmented files to check strengths and weaknesses of the different techniques in relation to fragmented and partial files.

Therefore, effective and efficient reassembling algorithms are needed to be investigated in order to improve the carver accuracy of the files carving and ultimately the recovery process. To overcome some of the limitations of the conventional image carving frameworks and algorithms, this research adopts machine learning and evolutionary algorithms in its main carving components of identification. Validation and reassembling. It is because the traditional algorithms are not able to overcome the Bifragmented and intertwined cases in the scan area of the images. It is due to the complexity of the problem search space and the solution derivation options which entails machine learning and evolutionary algorithms that have heuristic capabilities.

The Extreme Learning Machine (ELM) is a feedforward neural network with one hidden layer that its proposed by (Huang *et al*., 2015). The weights of the hidden layer are chosen randomly and never updated. The weights of the output layer are learned in one iteration. It is used to solve complex problems in the classification, prediction and clustering domain and some example are Roux (2008) and Zhang *al et*. (2016). In this work, the identification and validation techniques encompass the ELM

REFERENCES

Abdullah, N. A., Mohamad, K. M., Ibrahim, R., & Deris, M. M. (2016). X_myKarve: Non-Contiguous JPEG File Carver. *International Journal of Digital Crime and Forensics (IJDCF)*, 8(3), 63-84.

Abdullah, N. A. (2014). *An improved file carver of intertwined jpeg images using X_myKarve.* Doctoral thesis, Universiti Tun Hussein Onn Malaysia.

Abdullah, N. A., Ibrahim, R., & Mohamad, K. M. (2012). Cluster size determination using JPEG files. *In International Conference on Computational Science and Its Applications*. Springer, Berlin, Heidelberg. pp. 353-363.

Alherbawi, N., Shukur, Z., & Sulaiman, R. (2013). Systematic literature review on data carving in digital forensic. *Procedia Technology*, 11, pp. 86-92.

Alshammari, E., Al-Naymat, G., & Hadi, A. (2017). A New Technique for File Carving on Hadoop Ecosystem. *In New Trends in Computing Sciences, 2017 International Conference on*. pp. 72-77.

Amirani, M. C., Toorani, M., & Beheshti, A. (2008). A new approach to content-based file type detection. *in IEEE Symposium on Computers and Communications*. pp. 1103-1108.

Amirani, M. C., Toorani, M., & Mihandoost, S. (2013). Feature- based Type Identification of File Fragments. *Security and Communication Networks, 6(1)*, pp.115-128.

Aronson, L., & Van Den Bos, J. (2011). Towards an engineering approach to file carver construction. *In Computer Software and Applications Conference Workshops (COMPSACW), 2011 IEEE*, pp. 368-373.

Axelsson, S. (2010). The Normalised Compression Distance as a file fragment classifier. *digital investigation*, 7, pp. S24-S31.

Birmingham, B., Farrugia, R. A., & Vella, M. (2017). Using thumbnail affinity for fragmentation point detection of JPEG files. In Smart Technologies, IEEE *EUROCON 2017-17th International Conference on*. pp. 3-8.

Bodziak, W. J. (2017). Forensic footwear evidence. *CRC Press*.

Cai, J., Dawson, L., Javan, G. T., Özsoy, S., Quaak, F. C., & Ralebitso-Senior, T. K. (2018). From Experimental Work to Real Crime Scenes and the Courts. *In Forensic Ecogenomics*, pp. 177-209.

Calhoun, W. C., & Coles, D. (2008). Predicting the types of file fragments. *digital investigation*, 5, pp. S14-S20.

Carrier, B. (2005). File system forensic analysis. *Journal of Chemical Information and Modelling*, 53-9, pp. 1689-1699.

Cohen, M. I. (2007). Advanced carving techniques. *Digital Investigation*, 4(3-4), pp.119-128.

De Bock, J., & De Smet, P. (2016). JPGcarve: an advanced tool for automated recovery of fragmented JPEG files. *IEEE Transactions on Information Forensics and Security, 11(1)*, pp. 19-34.

Dewald, A., Luft, M., & Suleder, J. (2018). Incident Analysis and Forensics in Docker Environments.

Digital Forensic Research Workshop. (2006). *DFRWS 2006 Forensics Challenge Details*. Retrieved on July, 2016.

Digital Forensic Research Workshop. (2007). *DFRWS 2006 Forensics Challenge Details*. Retrieved on July, 2016.

Durmus, E., Korus, P., & Memon, N. (2019). Every Shred Helps: Assembling Evidence From Orphaned JPEG Fragments. *IEEE Transactions on Information Forensics and Security*, 14(9), 2372-2386.

Efthymia, T., & Ioannis, P. (2009). Automatic color based reassembly of fragmented images and paintings. *IEEE Transactions on Image Processing, 19(3),* pp. 680-690.

FICCI (2015), Pinkerton C&I India Ltd. *FICCI Indian Risk Survey 2015*.

Ganesh, V. (2017). Artificial Intelligence Applied to Computer Forensics. *International Journal*, 5(5), pp. 21-29.

Garfinkel, S. L. (2007). Carving contiguous and fragmented files with fast object validation. *digital investigation, 4*, pp. S2-S12.

Garfinkel, S. L. (2010). Digital Forensics Research: The Next 10 Years. *Digital Investigation, 7(1)*, pp. S64-S73.

Garfinkel, S. L. (2012). *File Carving*. < http://www.forensicswiki.org/wiki/File_Carving>.

Guo, H., & Xu, M. (2011). A method for recovering jpeg files based on thumbnail. *In Control, Automation and Systems Engineering, 2011 International Conference on*. pp. 1-4.

Hamilton, E. (1992). JPEG File Interchange Format Version 1.02. <http://www.w3.org/Graphics/JPEG/jfif3.pdf>.

Hand, S., Lin, Z., Gu, G., & Thuraisingham, B. (2012). Bin-Carver: Automatic recovery of binary executable files. *Digital Investigation*, 9, pp. S108-S117.

Hard drive physical sectors architecture and data reading process, ACE Data Group, <https://www.datarecovery.net/articles/hard-drive-sectordamage.aspx>.

Hilgert, J. N., Lambertz, M., Rybalka, M., & Schell, R. (2019). Syntactical Carving of PNGs and Automated Generation of Reproducible Datasets. Digital Investigation, 29, S22-S30.

Huang, G. B., Bai, Z., Kasun, L. L. C., & Vong, C. M. (2015). Local receptive fields based extreme learning machine. *IEEE Computational Intelligence Magazine*, 10(2), pp. 18-29.

Huang, J. D. *The JPEG Standard*. Graduate Institute of Communication Engineering National Taiwan University, 2006.

Karabiyik, U. *Building an intelligent assistant for digital forensics*. Doctoral thesis, The Florida State University, 2015.

Karresand, M., & Shahmehri, N. (2006). Oscar—file type identification of binary data in disk clusters and ram pages. *In IFIP International Information Security Conference*. Springer, Boston, MA, pp. 413-424.

Karresand, M., & Shahmehri, N. (2008). Reassembly of fragmented JPEG images containing restart markers. *Proceeding of the 2008 European Conference on Computer Network Defense*. pp. 25-32.

Kendall, K., Kornblum, J., & Mikus, N. (2011). *Foremost*. < http://foremost. sourceforge. Net>.

Kenney, E. L., & Gortmaker, S. L. (2017). United States adolescents' television, computer, videogame, smartphone, and tablet use: associations with sugary drinks, sleep, physical activity, and obesity. *The Journal of paediatrics, 182*, pp. 144-149.

Kloet, S. J. J. *Measuring and improving the quality of file carving methods.* Master's Thesis. Eindhoven University of Technology, 4-79, 2007.

Lee, I. H., Mahmood, M. T., Shim, S. O., & Choi, T. S. (2014). Optimizing image focus for 3D shape recovery through genetic algorithm. *Multimedia tools and applications*, 71(1), 247-262.

Lewis, A. B. *Reconstructing compressed photo and video data.* Doctoral thesis. University of Cambridge, Computer Laboratory, 2012.

Li, Q., Sahin, B., Chang, E. C., & Thing, V. L. (2011). Content based JPEG fragmentation point detection. *In Multimedia and Expo, 2011 IEEE International Conference on*. pp. 1-6.

Li, W. J., Wang, K., Stolfo, S. J. & Herzog, B. (2005). Fileprints: Identifying File Types by n-gram Analysis. Proceeding of the 2005 IEEE Workshop on Information Assurance and Security. West Point, New York. pp. 64-71.

McDaniel, M. B. *An algorithm for content-based automated file type recognition*. Master's thesis, James Madison University, Harrisonburg, United States, 2001.

McDaniel, M., & Heydari, M. H. (2003, January). Content based file type detection algorithms. In System Sciences, 2003. P*roceedings of the 36th Annual Hawaii International Conference on*. pp. 10-pp.

Memon, N., & Pal, A. (2006). Automated reassembly of file fragmented images using greedy algorithms. *IEEE Transactions on Image Processing*, 15(2), pp. 385-393.

Metz, J. & Mora, R. J. (2006). Analysis of 2006 DFRWS Forensic Carving Challenge.

Mikus, N. *An Analysis of Disc Carving Techniques.* Master's Thesis. Naval Postgraduate School; 2005.

Mohamad, K. M., & Deris, M. M. (2009). Visualization of JPEG metadata. *In International Visual Informatics Conference*. Springer, Berlin, Heidelberg, pp. 543-550.

Mohamad, K. M., Patel, A., & Deris, M. M. (2011). Carving JPEG Images and Thumbnails Using Image Pattern Matching. *Proceeding of the 2011 IEEE Symposium on Computers & Informatics*. pp. 78-83.

Mohamad, K. M., Patel, A., Herawan, T., & Deris, M. M. (2010). myKarve: JPEG image and thumbnail carver. *Journal of Digital Forensic Practice, 3(2-4)*, pp. 74-97.

Mohammed, M. A., Al-Khateeb, B., Rashid, A. N., Ibrahim, D. A., Ghani, M. K. A., & Mostafa, S. A. (2018). Neural network and multi-fractal dimension features

for breast cancer classification from ultrasound images. *Computers & Electrical Engineering*.

Mohammed, M. A., Ghani, M. K. A., Hamed, R. I., Mostafa, S. A., Ibrahim, D. A., Jameel, H. K., & Alallah, A. H. (2017). Solving vehicle routing problem by using improved K-nearest neighbour algorithm for best solution. *Journal of Computational Science*, 21, pp. 232-240.

Mostafa, S. A., Mustapha, A., Mohammed, M. A., Ahmad, M. S., & Mahmoud, M. A. (2018). A fuzzy logic control in adjustable autonomy of a multi-agent system for an automated elderly movement monitoring application. *International journal of medical informatics*, 112, pp. 173-184.

Nadeem Ashraf, M. *Forensic Multimedia File Carving*. Master's Thesis. Department of Computer and Systems Sciences Royal Institute of Technology, 2013.

Nguyen, C. *Computer Faults in JPEG Compression and Decompression Systems*. A proposal submitted in partial fulfillment of the requirements for the qualifying exam. Electrical and Computer Engineering University of California, 2002.

Pahade, R. K., Singh, B., & Singh, (2015). A Survey on Multimedia File Carving. *International Journal of Computer Science & Engineering Survey, Vol.6*, No.6

Pal, A., & Memon, N. (2009). The evolution of file carving. *IEEE signal processing magazine*, 26(2), pp. 59-71.

Pal, A., Sencar, H. T. & Memon, N. (2008). Detecting file fragmentation point using sequential hypothesis testing. *digital investigation*, 5, pp. S2-S13.

Pal, A., Shanmugasundram, K. & Memon, N. (2003). Automated Reassembly of Fragmented Images. *Proceeding of the 2003 Acoustics Speech and Signal Processing (ICASSP).* Vol. 1, pp. I-625.

Pal, S. K., & Wang, P. P. *Genetic algorithms for pattern recognition*. CRC press, 2017.

Panda, S. S., Jena, G., & Sahu, S. K. (2015). Image super resolution reconstruction using iterative adaptive regularization method and genetic algorithm. *In Computational Intelligence in Data Mining* (pp. 675-681).

Povar, D., & Bhadran, V. K. (2010). Forensic data carving. *In International Conference on Digital Forensics and Cyber Crime*. Springer, Berlin, Heidelberg, pp. 137-148.

Qiu, W., Zhu, R., Guo, J., Tang, X., Liu, B., & Huang, Z. (2014, November). A new approach to multimedia files carving. I*n Bioinformatics and Bioengineering, 2014 IEEE International Conference on.* pp. 105-110.

Quick, D., & Choo, K. K. R. (2014). Data reduction and data mining framework for digital forensic evidence: storage, intelligence, review and archive. C*rime and Criminal Justice*, pp. 1-11.

Rao, G. S. V. R. K., Jinka, P., Srinivasan, V., Selvaraj, R., Ramaswamy, S. K., & Maroo, D. (2015). System and method for automatically extracting multi-format data from documents and converting into XML. *Washington, DC: U.S.* Patent No. 9,158,744.

Razali, N. M., & Geraghty, J. (2011). Genetic algorithm performance with different selection strategies in solving TSP. *In Proceedings of the world congress on engineering, Vol. 2*. Hong Kong, pp. 1134-1139.

Richard III, G. G. & Roussev, V. (2005). Scalpel: A Frugal, High Performance File Carver. *Proceeding of the 2005 Digital Forensics Research Workshop*. New Orleans.

Richard, G., Roussev, V., & Marziale, L. (2007). In-place file carving. *In IFIP International Conference on Digital Forensics*. Springer, New York, NY, pp. 217-230.

Roux, B. *Reconstructing Textual File Fragments Using Unsupervised Machine Learning Techniques*. Master's Thesis. University of New Orleans, 2008.

Sencar, H. T. & Memon, N. (2009). Identification and Recovery of JPEG Files with Missing Fragments. *Digital Investigation*, 6, pp. S88-S98.

Serbes, A., & Durak, L. (2010). Optimum signal and image recovery by the method of alternating projections in fractional Fourier domains. *Communications in Nonlinear Science and Numerical Simulation*, 15(3), 675-689.

Shannon, C. E. (2001). A mathematical theory of communication. *ACM SIGMOBILE mobile computing and communications review*, 5(1), pp. 3-55.

Shannon, M. (2004). Forensic relative strength scoring: ASCII and entropy scoring. *International Journal of Digital Evidence*, 2(4), pp 1-19.

Singh, A., Jindal, N., & Singh, K. (2016). A review on digital image forensics. *International Conference on Signal Processing*. pp. 12-6.

Sorokin, A., & Makushenko, E. (2016). Identification of JPEG files fragments on digital media using binary patterns based on Huffman code table. *In Digital Information Processing, Data Mining, and Wireless Communications 2016 Third International Conference on*. pp. 137-141.

Sportiello, L. & Zanero, S. (2011). File Block Classification by Support Vector Machine. *Proceeding of Sixth International Conference on Availability, Reliability and Security*. Vienna. pp. 307-312.

Standard of Japan Electronics and Information Technology Industries Association (2002). *JEITA 2002 Exchangeable Image File Format for Digital Still Cameras Exif Version 2.1*, CP-3451.

Sun, Y., Zhang, X., Jian, M., Wang, S., Wu, Z., Su, Q., & Chen, B. (2018). An improved genetic algorithm for three-dimensional reconstruction from a single uniform texture image. *Soft Computing*, 22(2), 477-486.

Tang, Y., Fang, J., Chow, K. P., Yiu, S. M., Xu, J., Feng, B., Li, Qiong. & Han, Q. (2016). Recovery of heavily fragmented JPEG files. *Digital Investigation, 18*, pp. S108-S117.

The International Telegraph and Telephone Consultative Committee (1992). CCITT 1992 *Information Technology Digital Compression and Coding of Continuous Tone Still Image Requirements and Guideline. T.81*. 1992.

Uzun, E., & Sencar, H. T. (2015). Carving orphaned JPEG file fragments. *IEEE Transactions on Information Forensics and Security, 10(8)*, pp. 1549-1563.

Veenman, C. J. (2007). Statistical Disk Cluster Classification for File Carving. *The Third International Symposium on Information Assurance and Security*. Manchester. pp. 393-398.

Wallace, G. K. (1992). The JPEG still picture compression standard. *IEEE transactions on consumer electronics*, 38(1), pp. xviii-xxxiv.

Wang, Y. (2016). Motion Blurred Image Restoration Based on Improved Genetic Algorithm. Rev. *Téc. Ing. Univ. Zulia*, 39(5), 231-237.

Xu, M., & Dong, S. (2009). Reassembling the fragmented JPEG images based on sequential pixel prediction. *In Computer Network and Multimedia Technology, CNMT 2009. International Symposium on*, pp. 1-6.

Wu, X., Han, Q., Niu, X., Zhang, H., Yiu, S. M., & Fang, J. (2018). JPEG image width estimation for file carving. *IET Image Processing*, 12(7), 1245-1252.

Zhang, J., & Dong, Q. (2016). Efficient ID-based public auditing for the outsourced data in cloud storage. *Information Sciences*, 343, pp. 1-14.

Zhang, L., Zhang, D., & Tian, F. (2016). SVM and ELM: who wins? Object recognition with deep convolutional features from imagenet. *In Proceedings of ELM-2015 Volume 1*. Springer, Cham, pp. 249-263.