GRAPHICS PROCESSING UNIT BASED PARALLEL COPY MOVE IMAGE FORGERY DETECTION SCHEME

AHMAD UWAYS BIN ZULKURNAIN

A project report submitted in partial fulfilment of the
requirements for the award of the degree of
Master of Computer Science (Information Security)

Faculty of Computing
Universiti Teknologi Malaysia

JANUARY 2015

To my family, especially my wife, Noridah, for unending support she has given me and my son, Ihsan, for inspiring me to improve. They are my motivation.

ACKNOWLEDGEMENT

Thank you to Dr. Mohd Fo'ad Rohani for his guidance, support and understanding. He is extremely accommodating and sincere in helping me complete the project.

Thank you to my wife and parents for their support. Without them, I could not have completed this project.

Finally, thank you to all my course mates who have given me feedback and were gracious enough to share their knowledge which helped me successfully execute this project.

I would also like to thank the developers of the utmthesis LaTeX project for making the thesis writing process a lot easier for me. Thanks to them, I could focus on the content of the thesis, and not waste time with formatting issues. Those guys are awesome.

ABSTRACT

In digital image forensics, an important area of research is forgery detection. Copy-move forgery is a specific type of image tampering where a part of the image is copied and pasted on some other part of the same image. Currently, robust copy move image forgery detection techniques are complex and face the problem of high computation time. CPU based and partial GPU based versions of copy move image forgery detection schemes currently exist, but parallelization can be improved to further reducing computation time. In this project, a fully GPU based detection scheme was designed and developed to achieve improved performance. In addition, this project uses counting bloom filters instead of radix sort for detecting duplicated image regions. To compare counting bloom filters with radix sort for duplicate detection, a detection scheme which supports both techniques is developed. The effectiveness of counting bloom filter is tested for robustness against copy move image forgeries with added post-processing and geometric transformations. The developed GPU based scheme is five times faster than multi-threaded CPU implementations for the feature extraction process while counting bloom filters performed 18 times faster than radix sort in duplicate detection. The scheme also achieves 84% detection rate. No false positives were detected by the scheme.

ABSTRAK

Dalam forensik imej digital, salah satu bidang penting dalam penyelidikan adalah pengesanan pemalsuan. Pemalsuan secara salin dan tampal adalah sejenis teknik pengubahan imej tertentu di mana sebahagian daripada imej disalin dan dialihkan ke bahagian lain dalam imej yang sama. Kaedah mantap untuk mengesan pemalsuan secara salin dan tampal dalam imej digital yang kini wujud adalah kompleks dan menghadapi masalah masa pengiraan yang tinggi. Skim pengesanan berasaskan CPU sepenuhnya dan separa berasaskan GPU bagi pengesanan pemalsuan imej secara salin dan tampal telah dibangunkan, tetapi penyelarian masih boleh diperbaiki untuk mengurangkan masa pengiraan. Projek ini mereka bentuk dan membina sebuah skim pengesanan berasaskan GPU sepenuhnya untuk mencapai prestasi yang lebih baik. Di samping itu, projek ini menggunakan *counting bloom filter* sebagai alternatif kepada penyusunan radix untuk mengesan kawasan imej yang hampir sama. Untuk membandingkan *counting bloom filter* dengan penyusunan radix dalam proses mengenal pasti ciri imej berpadanan, satu skim pengesanan yang menyokong kedua-dua teknik dibangunkan. Keberkesanan *counting bloom filter* diuji untuk ketegapan terhadap pemalsuan imej secara salin dan tampal dengan penambahan pasca pemprosesan dan transformasi geometri kawasan imej yang disalin. Skim berasaskan GPU yang dibangunkan lima kali lebih pantas daripada pelaksanaan CPU untuk proses pengekstrakan ciri manakala *counting bloom filter* 18 kali lebih pantas daripada penyusunan radix dalam pengesanan pendua. Skim ini juga mencapai 84% kadar pengesanan. Tiada pengesanan palsu berlaku dengan skim tersebut.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AWGN	–	Additive White Gaussian Noise
CASIA	–	Institute of Automation, Chinese Academy of Sciences
CPU	–	Central Processing Unit
CUDA	–	Compute Unified Device Architecture
DCT	–	Discrete Cosine Transform
DWT	–	Discrete Wavelet Transform
DyWT	–	Dyadic Wavelet Transform
FFT	–	Fast Fourier Transform
FMT	–	Fourier-Mellin Transform
GPU	–	Graphics Processing Unit
I/O	–	Input/Output
IMD	–	Image Manipulation Dataset
IPC	–	Instructions Per Cycle
JPEG	–	Joint Photographic Experts Group
KPCA	–	Kernel Principal Component Analysis
LBP	–	Local Binary Pattern
LSD	–	Least Significant Digit
MIMD	–	Multiple Instruction, Multiple Data
MSD	–	Most Significant Digit
PCA	–	Principal Component Analysis
PCA-EVD	–	Principal Component Analysis-Eigen Value Decomposition
PCT	–	Principal Component Transformation
PHT	–	Polar Harmonic Transform
QCD	–	Quantized Coefficient Decomposition
SIFT	–	Scale Invariant Feature Transform
SVD	–	Singular Value Decomposition

LIST OF SYMBOLS

b	–	Block size
r	–	Radius
M	–	Image dimension
N	–	Alternate image dimension
$R_{detection}$	–	Detection rate
R_{false}	–	False detection rate
x	–	X-coordinate of image
y	–	Y-coordinate of image
p	–	Pixel
Y'	–	Luma channel
R'	–	Red channel
G'	–	Green channel
B'	–	Blue channel
G	–	DCT coefficient matrix
Q	–	Quantization matrix
B	–	Quantized coefficient matrix

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Introduction

Image forgery detection is needed to prevent alteration of images and restore some trust in digital images (Farid, 2009). It is applied in areas such as journalism, digital forensic science, and surveillance systems (Gupta, Saxena and Vasistha, 2013). The availability of powerful image processing and editing software makes it easy to create, alter, and manipulate digital images (Fridrich, Soukal and Lukáš, 2003). With that, the issue of verifying the authenticity and integrity of digital images is becoming increasingly important. There are two categories of image forgery detection techniques: active and passive. Active, also known as intrusive, detection techniques require a form of digital signature to be embedded in the image at the instance of its creation. However, not all digital devices are able to implant such signatures when capturing images (Muhammad *et al.*, 2011). On the other hand, passive, also referred to as non-intrusive or blind, approaches examine the image blindly without reliance on any embedded information. Although a passive approach has wider scope of usefulness, it is a computationally expensive process (Khan and Kulkarni, 2010).

Copy-move forgery is one of the tampering methods used to manipulate digital images. It is done by duplicating a region of the original image and pasting it onto another region of the same image. Various methods have been proposed to passively detect copy move image forgery. There are two main categories of approach for copy move image forgery detection: block based and keypoint based. Block based approaches identify image features based on local image regions whereas keypoint based approaches identify feature descriptions of objects within the image. This project will focus on block based approaches and the issues specific to this class of copy move image forgery detection.

The purpose of block based copy move image forgery detection is to avoid an exhaustive comparison of pixels, which is impractical for large images which are made up of millions of pixels. Segmenting the image into blocks reduces the number of features which need to be compared. In order to produce small but accurate representation of blocks, many feature extraction techniques have been proposed such as Discrete Cosine Transform (DCT) (Fridrich *et al.*, 2003), Principal Component Analysis (PCA) (Farid and Popescu, 2004), Discrete Wavelet Transform (DWT) (Bashar *et al.*, 2010), Singular Value Decomposition (SVD) (Ting and Rang-ding, 2009), and Fourier-Mellin Transform (FMT) (Bayram *et al.*, 2009). Some hybrids and sub-variations of the aforementioned techniques have also been proposed. Although many approaches have been proposed, there are still issues that need to be solved.

Before the issues that exist in block based approaches to copy move image forgery detection can be explored, the processes within a block based detection scheme must first be identified and explained. The common flow of block based copy move forgery detection can be seen in Figure 1.1. The input for the detection scheme is an image suspected to contain copy move forgery. Firstly, the image is segmented into overlapping blocks to separate the different image regions. The image region within each block is then goes through a feature extraction process which transforms raw pixel information into a set of image features. The resulting set of image features is then subjected to a similarity analysis process which identifies pairs of highly similar or identical image features. The final output of the detection scheme is a set of blocks suspected to be duplicates of one another.

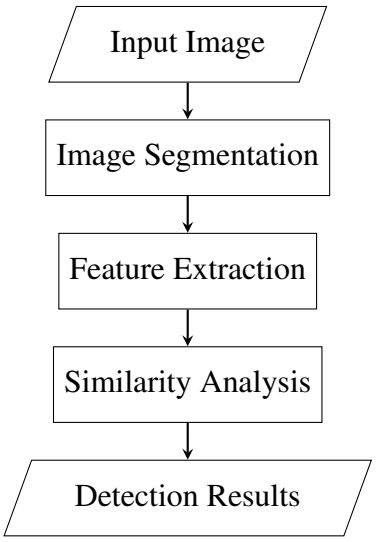

Figure 1.1: Copy Move Image Forgery Detection Process

A robust copy move image forgery detection scheme needs to be invariant to post-processing of the tampered image and geometric transformations of the the copy moved image regions. From Figure 1.1, feature extraction and similarity analysis are the most important processes and thus have been the focus of research in this field. As research progresses towards developing more robust methods of feature extraction, the algorithms being proposed have higher and higher time complexity. The complex transformations used to create feature vectors of image blocks require high computation time. The feature vectors produced during the feature extraction process are also complex and require some form of normalization or quantization before similarity analysis can be done, requiring extra rounds of calculations which increases computation time. As for the similarity analysis process, an efficient technique is needed to quickly identify duplicate image features within the set of obtained feature vectors.

1.2 Problem Background

Computation time is currently a serious issue in copy move image forgery detection, especially in real-time and near-real time applications. Resorting to less complex algorithms for feature extraction results in lower successful detection rate and higher false detection rate. These less complex algorithms are also less robust when confronted with post-processing of the tampered image and geometric transforms on the the copy moved portions of the image. Hence, it is not a viable solution because it decreases the overall success and accuracy of forgery detection. It is also possible to reduce the dimensions of an image before it is analyzed to gain a smaller search space. However, doing so causes data loss of image detail resulting in less accurate detection. One method of preserving the robustness of copy move image forgery detection while improving performance is through parallelization of the feature extraction process. Parallelization reduces the impact of high algorithmic time complexity and allows the total time required for forgery detection to scale with the hardware running the detection scheme.

While the techniques used for feature extraction in previous works have been various, the techniques used for duplicate feature matching have been extremely limited. A majority of proposed schemes in previous works have used lexicographical sorting of feature vector matrix to identify duplicated image regions. Improvement of computation time for lexicographical sorting has been achieved using radix sort. There has not been focus in current research on exploring more efficient methods to perform

block matching because lexicographical sorting has been proven to be effective without major drawbacks in terms of accuracy. However, the issue with lexicographical sorting is its computation time (Bayram, Sencar and Memon, 2009). Bayram, Sencar and Memon (2009) showed that identifying duplicate features using a counting bloom filter is much faster than lexicographical sorting. Despite this, there have not been any thorough investigations into how bloom filters perform in terms of resulting successful detection rate or false positive rate.

Currently, there exist some versions of copy move image forgery detection schemes using Central Processing Unit (CPU) based parallelization. Sridevi, Mala and Sandeep (2012) developed a parallelization algorithm based on Java threads executed on CPU. Modern CPUs offer high clockspeeds and Instructions Per Cycle (IPC) count and possess excellent pipelining capabilities to process a large number of threads simultaneously despite their limited number of physical processing cores. With the existence of Graphics Processing Unit (GPU) general purpose computing interfaces, it is possible to leverage a large number of GPU processing cores to execute tasks traditionally limited to CPUs. GPUs are designed specifically to process graphical data and are thus highly suited for image processing. GPUs have lower clockspeeds and IPC compared to CPUs but possess a large number of physical cores that can be used for processing. There has been use of GPUs to accelerate the radix sort of a copy move image forgery detection scheme, but there is currently no research into using GPU optimized counting bloom filters for copy move image forgery detection. There is also currently no algorithm to describe how copy move image forgery detection can be fully parallelized using a GPU. It is not possible to achieve the same optimization level by applying the same algorithms developed for CPU parallelization to GPU assisted computing because of issues such as host to device communication latency and separated memory. Thus, an algorithm must be designed specifically for the GPU in order to be optimal.

To evaluate the viability and performance of counting bloom filters in copy move image forgery detection schemes, a comparison must be made to the existing lexicographical sorting technique used for duplicate matching. Because a parallel radix sort has already been used on GPU to significantly speed up lexicographical sorting, a parallel bloom filter must be used for a fair comparison with radix sort. In addition, an equal comparison requires the feature extraction technique to remain constant and produces a feature set that is suitable for processing by both counting bloom filter and radix sort. Thus, a detection scheme which supports both radix sort and counting bloom filters needs to be developed. Performance evaluation of counting

bloom filters also needs to consider forgery detection under different scenarios such as post-processing of tampered image and geometric transformations of forged image regions.

1.3 Problem Statement

Robust feature extraction techniques used in block based copy move image forgery detection require a high computation time. There is a need for reduced computation time for detection schemes to be practical for use with large images and in real time environments. Utilization of GPU to compute the processes can highly parallelize the tasks involved to reduce computation time. Current CPU based algorithms that have been designed are not suitable to be directly adopted in a GPU based scheme. It must be determined how a parallel copy move image forgery detection scheme can be designed for use with a GPU. Counting bloom filter is faster than radix sort in processing feature vector matrix to identify duplicates. Evaluation of counting bloom filter in copy move image forgery detection scheme must be done to determine its effectiveness. To test the performance of counting bloom filter, a detection scheme which supports both radix sort and counting bloom filter needs to be developed.

1.4 Research Questions

The main research question to be answered in this study is:

"Are counting bloom filters an effective technique for duplicate matching in a GPU based copy move image forgery detection scheme?"

The supporting research questions are:

i. How can a GPU based copy move image forgery detection scheme using overlapping block technique be designed?

ii. Which feature extraction technique for image transformation to feature vector should be applied for evaluating the scheme that will be developed in this project?

iii. How can a radix sort be incorporated into a GPU based scheme?

iv. How can a counting bloom filter be developed for use with a GPU based detection scheme?

v. How does a parallel counting bloom filter perform in duplicate matching compare to radix sort?

1.5 Project Aim

The project aim is to improve the performance of copy move image forgery detection through GPU parallelization and evaluate alternative similarity analysis techniques by comparing the performance of similarity analysis using counting bloom filter with radix sort.

1.6 Project Objectives

The objectives for this research are as follows:

i. To design a parallel GPU based algorithm for copy move image forgery detection using overlapping block technique.

ii. To develop a copy move image forgery detection scheme based on the designed GPU algorithm which supports forgery detection using counting bloom filters and radix sort.

iii. To compare the performance of parallelized counting bloom filters with radix sort in performing forgery detection.

1.7 Project Scope

The scopes of this project are as follows:

i. Copy move image forgery detection can be done using either block based or keypoint based approaches. This project will only focus on block based

approaches for copy move image forgery detection.

ii. The parallel approach proposed will be specific to the type of GPU architecture and software framework for GPU programming in this project, which is based on Nvidia discrete graphics cards and Compute Unified Device Architecture (CUDA) platform.

iii. The project will not propose a new method for transformation of image data into feature representations used in the feature extraction stage. Instead, quantized Discrete Cosine Transform (DCT) will be adopted as a case study for parallelization of feature extraction.

iv. The images used for evaluating forgery detection will be generated from the Image Manipulation Dataset (IMD).

v. Methods have been proposed to perform duplicate matching using k-d tree, lexicographical sorting and bloom filter. In this project, only the radix sort implementation of lexicographical sorting and counting bloom filters will be evaluated.

vi. For evaluation of the parallelized counting bloom filters, comparison will only be made with radix sort for similarity analysis. The performance will be compared in terms of detection rate, false positive rate and speed.

1.8 Importance of Project

Currently, copy move image forgery detection is needed in many fields. Of those fields, journalism and digital forensics often deals with time sensitive cases. Therefore, a faster forgery detection system will benefit organizations in dealing with their cases in a timely manner. Existence of parallelized detection scheme can also make the more complex feature extraction techniques more practical by minimizing the computation time disadvantages. This project is also a step forward towards real-time or near real-time image forgery detection.

1.9 Thesis Organization

The thesis is organized in four chapters. This chapter gives and overview of the fundamentals of copy-move digital image forgery detection and overview of the current problems that aims to be solved by this project, specifically in regards to the

feature extraction and similarity analysis processes within a copy move image forgery detection scheme. The aim and objectives of the project are defined. Finally, The project scopes are defined and a the benefits of the project are explained.

In Chapter 2, previously done works in the field of copy move image forgery detection are reviewed and critically analyzed. An explanation of the current issues that exist within this area of research is presented. The study of past works also identifies the existing methods which meet the criteria for robustness and are suitable to be used with a GPU. The issues of performance with block matching techniques is also further explored. Finally, a study of GPU parallelization is done to help determine how it can be applied to copy move image forgery detection.

Methodology of the project is discussed in Chapter 3 which includes the experimental framework and other contributing factors of this project. In particular, dataset preparation, parameter selection and experimental process are discusses along with how the developed scheme can be evaluated. In Chapter 4, the design and implementation of the proposed scheme is shown along with details of the algorithms used. Chapter 5 presents the results and analysis obtained from the experiments carried out. Finally, Chapter 6 presents some concluding remarks and recommendation for future works.

REFERENCES

Abd-Elhafiez, W. M. and Gharibi, W. (2012). Color Image Compression Algorithm Based on the DCT Blocks. *International Journal of Computer Science Issues (IJCSI)*. 9(4).

Bashar, M., Noda, K., Ohnishi, N. and Mori, K. (2010). *Exploring Duplicated Regions in Natural Images*. doi:10.1109/TIP.2010.2046599.

Bayram, S., Sencar, H. T. and Memon, N. (2009). An efficient and robust method for detecting copy-move forgery. In *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*. IEEE, 1053–1056.

Bradski, G. and Kaehler, A. (2008). *Learning OpenCV: Computer vision with the OpenCV library*. " O'Reilly Media, Inc.".

Bravo-Solorio, S. and Nandi, A. K. (2011). Automated detection and localisation of duplicated regions affected by reflection, rotation and scaling in image forensics. *Signal Processing*. 91(8), 1759–1770.

Christlein, V., Riess, C., Jordan, J. and Angelopoulou, E. (2012). An evaluation of popular copy-move forgery detection approaches. *Information Forensics and Security, IEEE Transactions on*. 7(6), 1841–1854.

Farid, A. and Popescu, A. (2004). *Exposing digital forgeries by detecting duplicated image regions*. Technical report. Technical Report, TR2004-515, Department of Computer Science, Dartmouth College, Hanover, New Hampshire.

Farid, H. (2009). Image forgery detection. *Signal Processing Magazine, IEEE*. 26(2), 16–25.

Fridrich, J., Soukal, D. and Lukáš, J. (2003). Detection of copy-move forgery in digital images. In *Proceedings of Digital Forensic Research Workshop*. Citeseer.

Ghorbani, M., Firouzmand, M. and Faraahi, A. (2011). DWT-DCT (QCD) based copy-move image forgery detection. In *Systems, Signals and Image Processing (IWSSIP), 2011 18th International Conference on*. IEEE, 1–4.

Gupta, A., Saxena, N. and Vasistha, S. (2013). Detecting Copy Move Forgery In Digital Images. International Journal of Engineering Research and Applications.

International Journal of Engineering Research and Applications. 3(2), 94–97.

Jablin, T. B. (2013). *Automatic Parallelization for GPUs*. Ph.D. Thesis. Princeton University.

Jenkins, B. (2012). *Spookyhash: a 128-bit noncryptographic hash*. Retrievable at http://burtleburtle.net/bob/hash/spooky.html.

Khan, S. and Kulkarni, A. (2010). Reduced time complexity for detection of copy-move forgery using discrete wavelet transform. *International Journal of Computer Applications*. 6(7), 31–36.

Kirsch, A. and Mitzenmacher, M. (2006). Less hashing, same performance: Building a better bloom filter. In *In Proc. the 14th Annual European Symposium on Algorithms (ESA 2006*. 456–467.

Li, L., Li, S. and Wang, J. (2012). Copy-move forgery detection based on PHT. In *Information and Communication Technologies (WICT), 2012 World Congress on*. Oct. 1061–1065. doi:10.1109/WICT.2012.6409232.

Li, L., Li, S., Zhu, H., Chu, S.-C., Roddick, J. F. and Pan, J.-S. (2013). An Efficient Scheme for Detecting Copymove Forged Images by Local Binary Patterns. *Journal of Information Hiding and Multimedia Signal Processing*. 4(1), 46–56.

Lin, H.-J., Wang, C.-W., Kao, Y.-T. *et al.* (2009). Fast copy-move forgery detection. *WSEAS Transactions on Signal Processing*. 5(5), 188–197.

Luo, W., Huang, J. and Qiu, G. (2006). Robust detection of region-duplication forgery in digital image. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, vol. 4. IEEE, 746–749.

Mahdian, B. and Saic, S. (2007). Detection of copy–move forgery using a method based on blur moment invariants. *Forensic science international*. 171(2), 180–189.

Mohamadian, Z. and Pouyan, A. (2013). Detection of Duplication Forgery in Digital Images in Uniform and Non-uniform Regions. In *Computer Modelling and Simulation (UKSim), 2013 UKSim 15th International Conference on*. April. 455–460. doi:10.1109/UKSim.2013.94.

Muhammad, G., Hussain, M., Khawaji, K. and Bebis, G. (2011). Blind copy move image forgery detection using dyadic undecimated wavelet transform. In *Digital Signal Processing (DSP), 2011 17th International Conference on*. IEEE, 1–6.

Popescu, A. C. and Farid, H. (2005). Exposing digital forgeries in color filter array interpolated images. *IEEE Transactions on Signal Processing*. 53(10), 3948–3959.

Qiao, M., Sung, A., Liu, Q. and Ribeiro, B. (2011). A novel approach for detection of copy-move forgery. In *ADVCOMP 2011, The Fifth International Conference on*

Advanced Engineering Computing and Applications in Sciences. 44–47.

Salam, A. T. and Ghazali, B. S. (2013). STATE OF THE ART OF COPY-MOVE FORGERY DETECTION TECHNIQUES: A REVIEW. *International Journal of Computer Science Issues (IJCSI)*. 10(6).

Selvapeter, P. J. and Hordijk, W. (2009). Cellular automata for image noise filtering. In *Nature & Biologically Inspired Computing, 2009. NaBIC 2009. World Congress on*. IEEE, 193–197.

Singh, J. and Raman, B. (2012). A high performance copy-move image forgery detection scheme on GPU. In *Proceedings of the International Conference on Soft Computing for Problem Solving (SocProS 2011) December 20-22, 2011*. Springer, 239–246.

Sridevi, M., Mala, C. and Sandeep, S. (2012). Copy-move image forgery detection in a parallel environment. *Computer Science & Information Technology (CS & IT)*. 52, 19–29.

Ting, Z. and Rang-ding, W. (2009). Copy-Move Forgery Detection Based on SVD in Digital Image. In *Image and Signal Processing, 2009. CISP '09. 2nd International Congress on*. Oct. 1–5. doi:10.1109/CISP.2009.5301325.

Wang, J., Liu, G., Li, H., Dai, Y. and Wang, Z. (2009). Detection of Image Region Duplication Forgery Using Model with Circle Block. In *Multimedia Information Networking and Security, 2009. MINES '09. International Conference on*, vol. 1. Nov. 25–29. doi:10.1109/MINES.2009.142.

Welsh, T., Ashikhmin, M. and Mueller, K. (2002). Transferring Color to Greyscale Images. *ACM Trans. Graph.* 21(3), 277–280. ISSN 0730-0301. doi:10.1145/566654.566576. Retrievable at http://doi.acm.org/10.1145/566654.566576.

Zhang, J., Feng, Z. and Su, Y. (2008). A new approach for detecting Copy-Move forgery in digital images. In *Communication Systems, 2008. ICCS 2008. 11th IEEE Singapore International Conference on*. Nov. 362–366. doi:10.1109/ICCS.2008.4737205.

Zimba, M. and Xingming, S. (2011). DWT-PCA(EVD) Based Copy-move Image Forgery Detection. *International Journal of Digital Content Technology and its Applications*. 5(1).