# WATERMARKING TEXT DOCUMENT IMAGE USING PASCAL TRIANGLE APPROACH

SAMAR KAMIL KHUDHAIR

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

Faculty of Computing
UniversitiTeknologi Malaysia

JANUARY 2014

I cordially dedicate this thesis to biggest treasures of my life, my parents, who gave me their love, and also for their endless support and encouragement Mom and Dad

I love you for every second of my life

# ACKNOWLEDGEMENT

First and foremost, I offer my sincerest gratitude to my supervisor, Prof. Dr. Ghazali Bin Sulong, who has supported me throughout my thesis with his patience and knowledge whilst allowing me the room to work in my own way. I attribute the level of my master degree to his encouragement and effort and without him this thesis, too, would not have been completed or written. One simply could not wish for a better or friendlier supervisor.

Besides, I would like to thank the authority of UniversitiTeknologi Malaysia (UTM) for providing me with a good environment and facilities such as computer laboratory to computer laboratory to complete this project with software which I need during process.

Finally, I would also like to extend my thanks to my friends who have given me the encouragement and support when I needed it.

# ABSTRACT

Emergence of internet and other modern digital applications such as electronic publishing and digital library make it easy to reproduce and re-distribute digital contents thus give room to so many copyright violations including plagiarism and other illegal use of contents that need to be resolved. One way to prevent these illicit activities is to watermark the document before distribution. Thus, this research proposes a new text document image watermarking algorithm which emphasises on two most important measures, visual quality and robustness. In order to boost these measures, third least significant bit has been used for insertion. In addition, to further strengthen the technique, the Pascal Triangle is applied to determine the best position for embedding. Experimental results on the standard dataset have revealed that the proposed watermarking has achieved very encouraging results with PSNR and NCC averaged 54.95db and 0.98 respectively. In terms of robustness against adding noise attacks, the performance of the proposed technique, however, is less satisfactory

# ABSTRAK

Kemunculan internet dan aplikasi moden lainnya seperti penyiaran eletronik dan pustaka digital memudahkan pengeluaran dan pengagihan semula bahan digital memberi ruang kepada banyak pencabulan hak cipta termasuk plagerisma serta penggunaan bahan terlarang yang memerlukan penyelesaian. Salah satu cara untuk mengekang aktiviti haram ini adalah dengan mentera iarkan dokumen tersebut sebelum diagihkan. Oleh itu, kajian ini mencadangkan satu algoritma tera air dokumen imej baru yang memberi penekanan kepada dua ukuran penting iaitu kualiti visual dan keteguhan. Untuk meningkatkan uukuran tersebut, bit ketara ketiga terkecil digunakan untuk sisipan. Di samping itu, untuk mengukuhkan lagi teknik tersebut, tiga segi Pascal digunakan untuk menentukan kedudukan terbaik untuk proses pembenaman. Keputusan eksperimen menggunakan set data piawai telah mendedahkan bahawa tera air yang dicadangkan telah mencapai keputusan yang sangat menggalakkan dengan PSNR dan NCC masing-masing purata 54.95db dan 0.98. Dari segi keteguhan terhadap serangan penambahan hingar, prestasi teknik yang dicadangkan , bagaimanapun, adalah kurang memuaskan

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURE

# LIST OF ABBREVIATION

| | |
|---|---|
| HVS | Human Visual System |
| ISB | Intermediate Significant Bit |
| JPEG | Joint Photographic Expert Group |
| LSB | Least Significant Bit |
| NCC | Normalized Cross Correlation |
| PSNR | Peak Signal to Noise Ratio |
| TIFF | Tagged Image File Format |

# CHAPTER1

# INTRODUCTION

## 1.1 Introduction

Emergence of digital media technology allows the possibility of sending, receiving and sharing various file format of digital content. This simplifies and lowers the cost of modern communication and information exchange. With this advancement in technology a single file can be replicated into any number of identical copies within a short possible time. More so, the technology enables the file of any format to be sent and received across the internet. Many advantageous features of digital content make them acceptable and the primary form of keeping records in offices, schools and other organizations. Ubiquity of use makes the area one of hottest research field of the day.

On the other hand, the digital technology gave a way to many unwanted practices and frauds, including: access violation, copyright violation, and illegal contents alteration. To overcome such challenges, some protective measures like: access authenticity, privacy control, and copyright protection must be in practice. Problems of plagiarism and other related offences make the technology less beneficial.

There are many forms of digital media including video files, audio files image files, text files and multimedia files. The text files are the major and most

important part of digital information. The text files are the main target attackers being it the carrier of vital information like passwords and other key information of any organization. A protecting text files is critical and necessary in order to prevent the negative effect of illegal duplication and alteration of the contents. Many solutions were proposed to tackle the problem but among which digital watermarking techniques is the most promising (Topkara et al.,2007). Digital watermarking method, works by secretly inserting a hidden data such as copyright information into a text document. Separating the watermarking from the original text is very difficult since the digital watermarking is invisibly embedded into the actual information.

Digital watermark can simply be defined as a verification code that resides within the data invisibly and strong enough to withstand many form of attacks. The secret code is refer to as digital watermark image while the original document is refers as host or cover image. The technique of digital watermarking is very essential in the protection of intellectual property. The technique allows the genuine authors full control over their digital resources.

The technique of digital watermarking overcomes many challenges copyright violations including, illegal duplication or redistribution of digital content. There are many researches in the area of image file, audio files and video files but for the protection of text file the area is relatively new. Text file is the most common medium for today's information exchanges. Text files are the major component of newspaper, e-books, academic papers, journals and magazine. Similarly, to the general watermarking defined above, a digital text watermarking is a way of hiding digital watermark into a digital text by secretly storing some information that will help owner to tract his own text.

## 1.2 Problem Background

Previous research proposed many text watermarking techniques including: synonym based, syntactic tree based, text image, pre-supposition based, noun–verb based, word and sentence based, acronym based, and typo error based to mention a few. The entire works on digital text watermarking can be classified into three basic classes:

- An image based approach
- Syntactic approach
- Semantic approach

The image based approach secretly embeds watermark is in text image by modifying the inter line or words of the gaps in between lines and words (Macq and Vybomova,2007). The syntactic approach embed watermark into the syntactic structure of text and its bits by applying some digital transformation like: cleftingpassivization and topicalization. And the semantic approach incorporates some technique like: synonyms, acronyms (Cox et al., 2002) words spelling, pre-suppositions (Craver et al.,1998) and text meaning as text's semantics to enter the watermarking in text.

The binary text image algorithms for text watermarking are not strong enough to withstand attacks and easily by pass by re-typing. But syntactic approach mixes the algorithms with the natural language processing (NLP) therefore made it stronger. The algorithms are more protective but slow and inefficient syntactic analyzers, and it suffers the problem of syntactic and irreversible transformations.

The approach of semantics-based watermarking algorithms are language dependent therefore not frequently use. To overcome synonym substitution attacks, the synonym-based method has to be complimented with the powerful syntactic analyzer. Random semantic transformation cannot be applied on some documents

like: legal related, poetry and quotations due to their sensitivity properties and to protect the semantic connotation, text's value and meaning. Therefore, some portable and robust copyright protection methods that can support all kinds of texts and provide the required protection have to be developed. The new method has to address the different features of foreground and background and ensures the protection of all the text properties such as text meaning, word patterning, fluency, language rule and author writing style.

## 1.3 Problem Statement

The need to have an efficient text watermarking algorithm is imperative especially with respect to testing against attacks. Most of the recent works in the field are based on spatial domain technique (Ail et al.,2003). Spatial domain methods are more popular than the frequency domain methods, the method of spatial domain is more robust and the watermark is more hidden from the general view (Kostopoulos et al.,2003).

The technique can supports the embedding of large binary object (Pascal's Triangle) into an image, and also allows robustness to most common image processing tasks. The technique is more effective and produces very result if relatively large text file is use compare to the size watermark. Combining the self-similarity properties of Pascal's triangles together with effective embedding method can help to produce a very reliable watermarking strategy that can provide the required robustness for JPEG compression and other geometric transformations (Xeno et al.,3003).

## 1.4 Research Questions

The research will answer the following questions:

i. How to embed a watermark image in binary image without compromising the quality of the document?
ii. Where to place the watermark in the image?
iii. Is the technique robust enough for all attacks?

## 1.5 Aim of Study

The aim of this dissertation is to improve robustness and imperceptibility of text document watermarking using spatial domain by means of blank spaces in the words.

## 1.6 Objectives of the Study

This dissertation intends to achieve the following objectives:

1) To propose a new text document watermarking technique using Pascal triangle by means of black pixels in words.
2) To improve imperceptibility by employing third ISB method .
3) To evaluate the quality and robustness of the proposed watermarking technique using PSNR and NCC against some attacks such as Gaussian noise, salt and pepper noise and Poisson noise.

## 1.7 Scope of the Study

Scope of this research is mainly based on following items:

1) The suggested technique implemented by Borland Delphi language on windows XP environment.
2) The suggested technique applied on 512* 512 gray scale image of 10 text document as a host image and 64* 64 Binary image of UTM logo as a watermark image. The format of the both of them is Bmp.
3) The suggested technique uses the black pixels in words of the text document image for embedding watermark image.
4) The third ISB bit of black pixels in words is used to embed the watermark image in the text document image.

## 1.8 Significance of the study

The research will help in boosting the security of text documents, which include: newspapers, research papers, legal documents, letters and novels.

## 1.9 Organization of the Report

The entire report comprised of five chapters. The chapter one covers the introduction to the whole work which includes background of the study, the problem statements, aim and objectives and scope of the research. In chapter two, the related literature on previous research related to our work that is the aspect information hiding and digital watermarking will be discussed. The main emphases will be on text watermarking. Chapter three will discuss the detail methodology of the research

including the overall research frame work. The chapter four will give full result and the implementation procedure of the research. Also in the chapter analysis and evaluation of the final result will be conducted. Finally, the last chapter of the report will provide the summary of the entire work, conclusion and recommendation for future works.

# REFRENCES


Boneh, D., and Shaw, J., Collusion secure fingerprinting for digital data. *Proceedings of Crypto 95*, 1995, Springer LNCS 963, pp. 452–465.

Brassil, J., Low, S., Maxemchuk, N., and O' Gorman, L. Hiding information in document images. *Proceedings of the 29th Annual Conference on Information Sciences and Systems*, 1995, pp. 482-489.

Brassil, J., Low, S., Maxemchuk, N., and O' Gorman, L., Electronic marking and identification techniques to discourage document copying. *Proceedings of IEEE INFOCOM '94*, 1994 3, pp. 1278–1287.

Cox, I. J. Kilian, J., Leighton, T., and Shamoon, T., Secure spread spectrum watermarking for images, audio, and video. *Proceedings of the 1996 IEEE International Conference on Image Processing*, 1996, 3, pp. 243–256.

Craver, S., Memon, N., Yeo, B.-L., and Yeung, M. M., On the invertibility of invisible watermarking techniques. *Proceedings of the 1997 IEEE International Conference on Image Processing*, 1997, 1, pp. 540–543.

F. Y. Shih, S. Y. T. Wu, Combinational Image Watermarking in the Spatial and Frequency Domains, *Elsevier, Pattern Recognition* 36 (2003), pages 957-968.

Flikkema, P. G., Spread-spectrum techniques for wireless communications. *IEEE Signal Processing Magazine*, 1997, 14, 26–36.

G.-J. Yu, C.-S.Lu, H.-Y. Mark Liao, A Message based Cocktail Watermarking System, *Elsevier, Pattern Recognition* 36 (2003), pages 969-975.

Hartung, F., and Girod, B., Digital watermarking of MPEG-2 coded video in the bit stream domain. *Proceedings of the 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1997, 4, pp. 2621–2624.

Hartung, F., and Girod, B., Digital Watermarking of Raw and Compressed Video. *Proceedings of the European EOS/SPIE Symposium on Advanced Imaging and Network Technologies*, Berlin, Germany, Oct. 1996.

Hartung, F., and Girod, B., Watermarking of uncompressed and compressed video. *Signal Processing*, 1998, 66, 283–301.

Hern´andez, J. R., P´erez-Gonz´alez, F., and Rodr´guez, J. M., The impact of channel coding on the performance of spatial watermarking for copyright protection, *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1998. 11.

I. J. Cox, J. Kilian, T. Leighton and T. Shamoon, A Secure, Robust Watermark for Multimedia, *First Workshop on Information Hiding*, Newton Institute, Univ. of Cambridge, May 1996.

ISO/IEC 13818-2, *Generic Coding of Moving Pictures and Associated Audio, Recommendation H.262 (MPEG-2)*, International Standard, 1995.

J.Fridrich and M.Goljan, Protection of Digital Images using Self Embedding, *Symposium on Content Security and Data Hiding in Digital Media*, Newark, NJ, USA, May 1999.

Jerry Gao, Jacky Cai, K.P. & Shim, and S., 2005. A Wireless Payment System.*Institute of Electrical and Electronics Engineers*, (San Jose State University,USA).

Kutter, M., F. Jordan, and F. Bossen, Digital Signature of Color Images Using Amplitude Modulation, *in Proceedings of the SPIE, Storage and Retrieval for Image and Video Databases V*, 1997, pp. 518-526.

Low, S., Maxemchuk, N., Brassil, J., and O' Gorman, L., Document marking and identification using both line and word shifting.*Proceedings of IEEE INFOCOM '95*, 1995.

M.Yeung and F.Mintzer, “An Invisible Watermarking Technique for Image Verification”, Proc. ICIP ’97, Santa Barbara, California, at.1997.

O Ruanaidh, J. J. K., and Pun, T., Rotation, scale, and translation invariant digital image watermarking.*Proceedings of the 1997 IEEE International Conference on Image Processing*, 1997, 1, pp. 536–539.

O Ruanaidh, J. J. K., Dowling,W. J., and Boland, F. M., Watermarking digital images for copyright protection. *IEE Proceedings on Vision and Image Signal Processing*, 1996, 143, 250–256.

Petitcolas, F. A. P., Anderson, R. J., and Kuhn, M. G., Attacks on copyright marking systems.*Proceedings of the Second Workshop on Information Hiding*, Portland, Oregon, USA, Apr. 1998.

Petitcolas, F. A. P., *StirMark*, vers. 1.0. URL http://www.cl.cam.ac.uk/ fapp2/watermarking/image watermarking/stirmark/, 1998.

Pickholtz, R. L., Schilling, D. L., and Milstein, L. B., Theory of spread-spectrum communications—A tutorial.*IEEE Transactions on Communications*, 1982, COM- 30, 855–884.

Piva, A., Barni, M., Bartolini, F., and Cappellini,V., DCT-based watermark recovering without resorting to the uncorrupted original image. *Proceedings of the 1997 IEEE International Conference on Image Processing*, 1997, 1, pp. 520–523.

Qiao, L., and Nahrstedt, K., Watermarking schemes and protocols for protecting rightful ownership and customer' s right. Submitted to *Academic Press Journal of Visual Communication and Image Representation*, 1998.

R. G. van Schyndel, A. Z. Tirkel, C. F. Osborne, A Digital Watermark, *Proc. IEEE International Conference on Image Processing, ICIP-94*, 1994, Vol.2, pp.86-90.

S. Armeni, D. Christodoulakis, I. Kostopoulos, Y. Stamatiou, M. Xenos, A Transparent Watermarking Method for Color Images, *First IEEE Balcan Conference on Signal Processing, Communications, Circuits, and Systems*, June 2000, Istanbul, Turkey.

S. Wolfram Geometry of Binomial Coefficients, *American Mathematical Monthly*, Vol. 91, pages 566-571, November 1984.

S.A.M Gilani, I. Kostopoulos, A.N. Skodras, Adaptive Color Image Watermarking, *14th IEEE International Conference on Digital Signal Processing*, 1-3 July 2002,Santorini, Greece.

Seyed Ail (2012*).” An Improved Text Document Image Watermarking Algorithm Based On White Spaces In – Between Words*”, Master Computing, University Technology Malaysia.

Smith, J. R., and Comiskey, B. O., Modulation and information hiding in images. *Proceedings of the First Information HidingWorkshop*, Cambridge, U. K., May 1996.

Stefan Katzenbeisser, Fabien A.P. Petitcolas, *Information Hiding Techniques for Steganography and Digital Watermarking*, Artech House, 2000.

Stone, H. S., Analysis of attacks on image watermarks with randomized coefficients. NEC Technical Report, May 1996.

Swanson, M. D., Zhu, B., Chau, B., and Tewfik, A. H., Muster solution video watermarking using perceptual models and scene segmentation. *Proceedings of the 1997 IEEE International Conference on Image Processing*, 1997, 2, pp. 558-561.

Van Schyndel, R. G., Tirkel, A. Z., and Osborne, C. F., A digital watermark. *Proceedings of the 1994 IEEE International Conference on Image Processing*, 1994, 2, pp. 86–89.

Wolfgang, R. B., and Delp, E. J., A watermarking technique for digital imagery: Further studies. *Proceedings of the International Conference on Imaging Science, Systems, and Technology*, 1997, Las Vegas, pp. 279–287.

Xia, X.-G., Boncelet, C. G., and Arce, G. R., A multiresolution watermark for digital images, *Proceedings of the 1997 IEEE InternationalConference on Image Processing*, 1997, 1, pp. 548–551.

Zeng, W., and and Liu, B., On resolving rightful ownerships of digital images by invisible watermarks. *Proceedings of the 1997 IEEE International Conference on Image Processing*, 1997, 1, pp. 552–555.