



Faculty of Information and Communication Technology

AL-MUS'HAF IMAGE BINARIZATION BY USING THREADS

Mustafa Sabah Ali

Master of Computer Science (Internetworking Technology)

2017

**THE EFFECTS OF INFORMATION TECHNOLOGY INFRASTRUCTURE
ON SMEs PERFORMANCE IN IRAQ**

HAYDER ADIL ABDUL RAHEEM

**A thesis submitted
in fulfillment of the requirements for the degree of Master of Business Administration in
Technology and Innovation Management**

Faculty of Technology Management and Technopreneurship

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2017

DECLARATION

I declare that this research project entitled "The Effects of Information Technology Infrastructures on SMEs Performance In Iraq" is the result of my own research except as cited in the references. The research project has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature :

Name :

Date :

APPROVAL

I hereby declare that I have read this research Project and in my opinion this research Project is sufficient in terms of scope and quality for the award of Master of Business Administration in Technology and Innovation Management.

Signature :.....

Supervisor Name :.....

Date :.....

DEDICATION

This Research Project is exclusively dedicated to my beloved family members especially my father Adil Abdul Raheem and my mother Majida Ibrahim and my sisters for their love, endless support, and prayers throughout of master journey. Special gratitude goes to my friends and my cousin Ali Lateef for unlimited supporting in the distribution of my research project survey in Iraq.

ABSTRACT

The development of micro, small and medium-sized enterprises (MSMEs) has become an alternative strategy to expand the Iraq economy especially after oil price crisis. It also indorse the formation of more creative and decent jobs for women and men maintainable enterprise development, accommodating companies of all sizes across many segments, is a main element to successful private segment improvement. Recently, information technology recognized as a crucial development technology tool that boosts the performance of corporations and economic growth. Studies from various disciplines investigated the importance of information technology to firms' performance leaving a large body of literature. In addition, the factors towards information technology acceptance have been investigated and explored and very public theories and models such as UTAUT and TAM already emerged. However, the processes by which IT resources interact with other human and organizational resources, as well as the nature of these resources, have hardly been studied. In addition, the success of information technology implementation depends on several factors such as flexible infrastructure, compatibility and the labour capability to cope with new technology and eligible human resources who can easily absorb and adapt the IT functions in their duties. The absorptive capacity of the firms plays an important role in accelerating the match between information technology and firm performance. The absorptive capacity refers to the capability of to absorb the knowledge and implement the technology. This research was conducted to examine the relationship between information technology infrastructure and SMEs performance. It also scrutinises the mediating effect of absorptive capacity in the relationship between information technology infrastructure and SMEs performance in Iraq. The quantitative methodology was conducted and survey was used conducted to collect data from 162 small medium enterprises. The multiple regression analysis was used to analysis the data. The finding of this research emphases that IT infrastructure (IT connectivity, modularity and personal) has positive effect on firm performance. The outcome of multiple regression also shows that absorptive capacity has a mediating effect on the relationship between information technology infrastructure and SMEs performance

ABSTRACT

Segmentasi imej halaman Al-mus'haf adalah ukuran yang sangat sensitive. Hal ini disebabkan Al-mus'haf ialah kitab Allah swt, dan sebarang kesilapan akan mempengaruhi kesucian Al-mus'haf. Penyataan masalah dalam penyelidikan ini adalah bahawa sebarang langkah segmentasi imej memerlukan pengurusan dengan tahap piksel, dimana ia membawakan masa pemprosesan yang lama, dan juga menjadikan sebarang algoritma gagal dalam menghadapi pelaksanaan yang sebenar. Objectf penyelidikan ini adalah untuk mengurangkan masa pelaksanaan mewujudkan imej binari, di mana berdasarkan nilai ambang objek-objek Al-mus'haf seperti teks dan sempadan diwakili oleh angka 0 dan seluruh imej diwakili dengan angka 1. Sebaliknya, matlamat kedua penyelidikan ini menunjukkan penanguhan masa pelaksanaan antara menggunakan kaedah pelaksanaan selari berdasarkan teknik benang java dan tanpa menggunakannya. Metodologi penyelidikan ini menggunakan pelaksanaan selari dengan memanfaatkan teknik multitask CPU melalui penggunaan benang, untuk meningkatkan masa pelaksanaan. Pada pengakhiran ujian, hasil menunjukkan bahawa masa pelaksanaan untuk mencipta imej binari dapat dikurangkan dengan ketara, dengan menggunakan 10 benang peratusan peningkatan rata-rata lebih dari 90%. Kesimpulannya, pelaksanaan selari mempunyai peranan penting dalam jangka masa untuk memperbaiki masa pelaksanaan, dan ia agak disyorkan untuk sebarang ukuran pemprosesan imej

ACKNOWLEDGEMENTS

Alhamdulillah, Thanks to ALLAH the Almighty because of his blessing enable me to submit this research project. In the course of completing this research, I am interacting with many people that made me in debt with them. In particular, I express my sincere appreciation to my main thesis supervisor, Prof. Madya Dr. Md. Nor Hayati Bin Tahir for his thrust, compassion, valuable direction and unwavering supports thought out this academic journey. I also would like to thank many staffs at Faculty of Technology management and Technopreneurship (FPTT) for their enormous administrative and dedicated assistance.

Millions of Thanks also go to my friends and my cousin Ali Lateef for unlimited supporting distribution of my project survey. I would like to thanks the social medium as such Doctorates Support Group (DSG) because of immediate response and trusted information. This research project would not be realized without the facilitation of the role played by the specific actor's en route accomplishment of this research project.

Last but not least, thanks to my family especially to my father Adil Abdul Raheem and my mother Majida Ibrahim, and my sisters for the continuous support, beliefs and for implausible encouragement during the hardest of times.

TABLE OF CONTENTS

	PAGE
DECLARATION	
APPROVAL	
DEDICATION	
ABSTRACT	i
ABSTRAK	ii
ACKNOWLEDGEME	iii
TABLE OF CONTENTS	iv
LIST OF TABLES	v
LIST OF FIGURES	vi
CHAPTER	
1. INTRODUCTION	1
1.1 Research background	3
1.3 Problem Statement	3
1.4 Research Questions	4
1.5 Research Objectives	4
1.6 Project Significant	4
1.7 The Scope of Research	5
1.8 Expected Outcomes	5
1.9 Conclusion	5
2. LITERATURE REVIEW	7
2.1 Introduction	7
2.2 Connected-component labeling	8
2.3 Thresholding	10
2.3.1 Global Thresholding	11
2.3.2 Local Thresholding	12
2.4 Binarization	13
2.5 The Threads	13
2.5.1 The Needs of the Threads in the Image Processing	14
2.5.2 Why the parallel execution by using CPU	16
2.6 Al-Mus'haf Image Processing	16
2.7 Digital Documentation	17
2.8 Conclusion	19
3. RESEARCH METHODOLOGY	20
3.1 Introduction	20
3.2 Research Framework	20
3.2.1 Investigation Phase	20
3.2.2 Implementation Phase	21
3.3 Task framework	22
3.4 Experimental Test Framework	23
3.5 Research Tools	24
3.6 Conclusion	25
4. IMPLEMENTATION	26

4.1	Introduction	26
4.2	Data Collection	26
4.3	Implementation process	27
4.3.1	Threshold Detection	29
4.3.2	The Binarization	31
4.4	Conclusion	34
5.	RESULT AND TESTING	35
5.1	Introduction	35
5.2	Testing	35
5.3	Comparison	37
5.3.1	Comparison Result	39
5.4	Conclusion	43
6.	CONCLUSION AND FUTURE WORK	44
6.1	Introduction	44
6.2	Summary	44
6.3	Future Works / Further Research	46
	REFERENCES	47

LIST OF TABLES

TABLE	TITLE	PAGE
2.1	comparing work	18
3.1	Tools and programming languages that utilized in this research	24
4.1	discerption of the data set that used in this study	27
5.1	Hardware Description	36
5.2	The implementation time of creating the binary image	39
5.3	Improvement rate between using and without using threads	41

LIST OF FIGURES

FIGURE	TITLE	PAGE
1.1	Example of Al-mus'haf image	2
2.1	Connected-component labeling task	9
2.2	Describe thresholding step	10
2.3	Image of Lena Söderberg used in many image processing experiments	15
2.4	Example of an image with noisy black border and noisy text region	17
3.1	Framework of the binary image creation for Al-mus'haf image	23
4.1	Creating the binary image of Al-mus'haf by using threads	28
4.2	Image of Al-mus'haf divided horizontally	32
4.3	Gathering the result of each individual threads	33
4.4	The implementation of the individual threads	34
5.1	Result of Al-mus'haf image binarization	36
5.2	User interface of selecting the image	38
5.3	User interface of showing the result of the implementation	39
5.2	User interface of selecting the image	38

CHAPTER 1

INTRODUCTION

1.1 Introduction

Nowadays, the image processing considered one of the most significant measures in computer science area(Kumar, 2004). The image processing is general measure has sub-sections such as, character recognition and image enhancement. The importance of image enhancement comes, due to the fact(Kumar, 2004), that most of the information and the data in the world are documented as digital materials. Where sometimes images obtained from satellites and conventional and digital cameras lack in contrast and brightness because of the limitations of imaging sub systems and illumination conditions while capturing image(Kumar, 1995). The images may have different types of noise, in image enhancement; the goal is to accentuate certain image features for subsequent analysis or for image display.

Currently, we lived in the world where high performance personal computers are widespread almost everywhere, therefore the image processing applications are become very popular due to the availability of the performance requirements, such as high speed processors and large memory size. Every single method in image processing has standard phases. The first phase is always the data collection, where it's usually be the images that needs to implement the method on them, where in this research most of the data will be images of

Al-mus'haf pages, as shown in Figure 1.1. Some of the methods in image processing have preprocessing phase, classification, and post processing.



Figure 1.1 Example of Al-mus'haf image

The scope of this research focusing on the image processing for Al-mus'haf pages, where Al-mus'haf is considered one of the most important book in the Muslim communities. Al-mus'haf image processing is very sensitive issue, because Al-mus'haf is the book of Allah swt, therefore as especially in Al-mus'haf image segmentation measure should be studied carefully, due to any incorrect segmentation will affect the holiness of the book.

Further, Al-mus'haf comparatively considered a large book, wherefore the implementation time should has a high priority as well as possible, where any algorithm that doesn't take the implementation time into the consideration as a main factor will fail to face the real implementation requirements.

Therefore this research suggests that the threads technology can be used to reduce the processing time of binary image creation for Al-mus'haf pages. The binary image is

considered a preprocessing phase in the image segmentation that aimed to determine the background and the foreground in the images. In the research, the foreground is the objects in Al-mus'haf images, which is the text and the frame.

The threads is the technique that allows to the programmers to take the advantage of multiprocessor systems. Where, this research used the threads to enable the parallel execution of binary image creation for Al-mus'haf pages.

1.2 Research Background:

As mentioned, the digital copy of Al-mus'haf become very popular in the Muslims communities, due to the advantage of the digital copies, such as searchable text, easy access, cost savings and the security. Thus, the needs of improving the techniques to deal with Al-mus'haf digital copies become more interested issue. Since, Al-mus'haf is the book of Allah swt, therefore dealing with this measure needs to be a very precis.

1.3 Problem Statement:

The image processing measure deal with the pixel level, where any algorithm related to the image processing measure needs to check every single pixel in the image. That is lead to a long time of processing, where Al-mus'haf book consists of 114 Surah, distributed on 30 chapters, where the number of pages are ranging between 600 to 800 pages, based on the publishers. Accordingly, any applications of Al-mus'haf segmentation will be useless, if they aren't taking into the consideration a time implementation factor.

1.4 Research Questions

- How the binary image process for Al-Mus'haf happen in image processing domain?
- How the parallel execution affects on the implementation time of creating the binary image?

1.5 Research Objective

- To propose a method for creating the binary image by using threads.
- To assess the performance between creating the binary image with and without using threads.

1.6 Project Significant

Al-mus'haf is one of the most important book to the Muslims. In this research, the binary image creation of Al-mus'haf will be proposed. The result can be used in several phases of Al-mus'haf image segmentation, where the binary image as a result in this research considered a preprocessing phase to identify the elements in Al-mus'haf image, for the purpose of, remove the Illumination, which is the border of Al-mus'haf image.

1.7 The Scope of Research

The image segmentation measure has several phases, this research is focused on a binary image creation. The main data collection of this research is several types of Al-mus'haf images. The scope of this research is focused on exploit the ability of the multitask in the

current processors to enhance the implementation time of create the binary image for Al-mus'haf pages.

- i. Focus on images from mus'haf Al-Quran.
- ii. Computers without NVidia graphic card.
- iii. Focus on the execution time and not the accuracy.
- iv. Processed images are in binarization.
- v. Stage involved are pre-processing and binarization only.
- vi. The result of this research study will be use in Connected-component labeling algorithm ,where the binary image will consider as a preprocessing phase in the image segmentation.

1.8 Expected Outcomes

The main expected outcome of this research is an algorithm that able to create the binary image of Al-mus'haf pages under the parallel execution, for the purpose of obtain the minimum implementation time.

1.9 Conclusion

This thesis addressed the issue of image segmentation for sensitive digital of Al-mus'haf pages image. We proposed that, to reduce the implementation time of the binary image creation by utilize the parallel execution of Al-mus'haf images depends on the threads technique, which is considered one of the most powerful method that provided by Java programming language.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Since the most of the data and the information are documented as digital materials, therefore the image processing is considered one of the most important measure that needs a carefully studying. One of the purposes of image processing is to identify and recognize the shape of an object on a digital image(Rakhmadi, 2010), as well as the image processing of Al-mus'haf considered a sensitive issue require a high priority in terms of the processing, due to Al-mus'haf is a book of God swt.

The Connected-component labeling is an important step in the measure of the image segmentation. The fundamental processing for the CCL algorithm is using the approach of thresholding to detect the connected pixels that's belong to the same region(Walczyk, Armitage and Binnie, 2010).

Since every single image processing methods such as The Connected-component labeling algorithm require to check every single pixels in image ,where as a real implantation, that's led to a very large number from the processes, here we will face the problem of long time implantation for each image that require to processing(Chabrier, Laurent and Emile,

2005). One of the most powerful technique that's we introduced in this project, which is the parallelism processing.

By divide Al-mus'haf image into several sections, where here the role of threads methods is comes. Where the threads which is the technology that's allow us to exploit the ability of multi-tasking for the processor, by this technique we will execute the divided image's sections in parallel way, to create the a binary array represent Al-mus'haf image. The purpose from the binary array is use it in the image segmentation, where its considered one of the most common step in the measure of image processing.

2.2 Connected-component labeling

The labeling of the connected components considered one of the most important step in the measure of object recognition, which is a process for identifying a specific object in a digital image, where the detected pixels are grouped together and indexed with a unique label. The goal of connected component labeling is local connectivity information (Oliveira and Lotufo, 2010). For every pixel this is the information about which neighbored pixels are connected. In the case of a binary image, two neighboring pixels are called connected, if they are both white or black(Dillencourt, Samet and Tamminen, 1992), but in case if there are multimodal labeling, which means two neighboring pixels have the freedom to be connected or not. The two arbitrary pixels are considered connected, in the condition if there is a path between neighboring connected pixels.

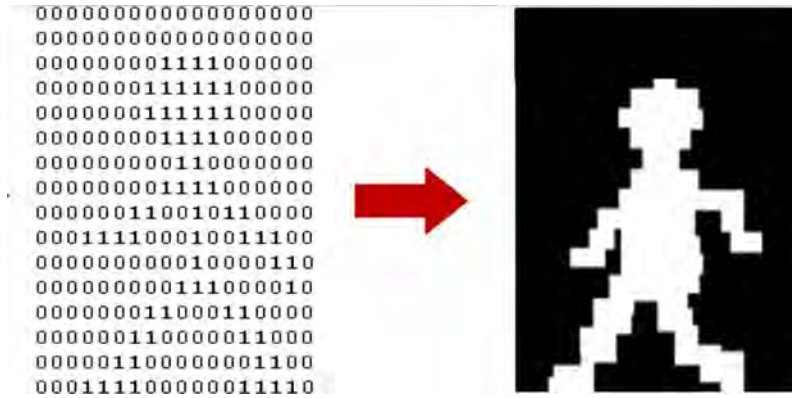


Figure 2.1 Connected-component labeling task

The original algorithm was developed by Rosenfeld and Pfaltz in 1966. It performs two passes through the image (Rakhmadi, 2010), in the first pass, the image is processed from left to right and top to bottom to generate labels for each pixel and all of the equivalent labels are stored in a pair of arrays. In the second pass, each label is replaced by the label assigned to its equivalence class.

In this project The technique of Connected-component labeling is using to represent the text of Al-mus'haf images , where the contents of Al-mus'haf image like the texts and the frame of the images are detecting by using the technique of image threshold.

connected component labeling algorithms can be classified in to four classes as (Walczyk, Armitage and Binnie, 2010), The first Method deal with repeated forward and backward passes over data. The second Method is two passes over data. These algorithms maintain a separate vector or an array to store equivalence information of labels and use search algorithms to resolve equivalences. The third Method using hierarchical tree structure to represent data. The last one is Methods using parallel algorithms. As these algorithms are used for binary images, this section considers pixel value 0 as belongs to background and pixel value 1 as belongs to an object of interest in a binary mages.

2.3 Thresholding

As mentioned, Segmentation involves separating an image into regions (or their contours) corresponding to objects. We usually try to segment regions by identifying common properties, which in this research represent Al-mus'haf images. Or similarly, we identify contours by identifying differences between regions. The simplest property that pixels in a region can share is intensity. So, a natural way to segment such regions is through thresholding, the separation of light and dark regions. Thresholding creates binary images from grey-level ones by turning all pixels below some threshold to zero and all pixels about that threshold to one.



Figure 2.2 Describe thresholding step

The major problem with thresholding is that we consider only the intensity, not any relationships between the pixels. The simplest (and most common) way to threshold an image is to set all pixels below a certain grey level to black, and to clear all others to white, but there is no guarantee that the pixels identified by the thresholding process are contiguous. We can easily include extraneous pixels that aren't part of the desired region, and we can just as easily miss isolated pixels within the region (especially near the boundaries of the region). These effects get worse as the noise gets worse, simply because it's more likely that a pixels intensity doesn't represent the normal intensity in the region. When we use thresholding, we typically have to play with it, sometimes losing too much of the region and sometimes getting too many extraneous background pixels.

On the other hand, The major problem with thresholding is that we consider only the intensity, not any relationships between the pixels. There is no guarantee that the pixels identified by the thresholding process are contiguous. We can easily include extraneous pixels that aren't part of the desired region, and we can just as easily miss isolated pixels within the region (especially near the boundaries of the region). These effects get worse as the noise gets worse, simply because it's more likely that a pixels intensity doesn't represent the normal intensity in the region.

When we use thresholding, we typically have to play with it, sometimes losing too much of the region and sometimes getting too many extraneous background pixels. Shadows of objects in the image are also a real pain not just where they fall across another object but where they mistakenly get included as part of a dark object on a light background.

2.3.1 Global Thresholding

In a way, thresholding can be seen as an extreme form of contrast enhancement, or making light pixels lighter and dark pixels darker. The simplest (and most common) way to threshold an image is to set all pixels below a certain grey level to black, and to clear all others to white. The question then is how to select this grey-level. One possibility is to pick the center of the range of possible values, so if pixels are eight bits deep (ranging from 0 to 255), for example, then 128 would be selected(Chaubey, 2016). This approach works well if all the "dark" pixels of the image really do have values under 128, and light pixels have values above 128, but if the image is over or under exposed, then the result might be all white or black. It is better to look at the range of actual values instead of possible values to

determine the threshold(Chaubey, 2016). The maximum and minimum values for each pixel in the image can be found, then the midpoint used as the threshold.

As even better way to select the threshold is not just to look at the range of actual values, but also their distribution. If, for example, you expect the image to be of black line art, or text on a white background then you expect that most pixels will be the intensity of the background and a smaller but significant proportion will be of the dark ink.

2.3.2 Local Thresholding

Another problem with global thresholding is that changes in illumination across the scene may cause some parts to be brighter (in the light) and some parts darker (in shadow) in ways that have nothing to do with the objects in the image(Singh and James, 2012). We can deal, at least in part, with such uneven illumination by determining thresholds locally. That is, instead of having a single global threshold, we allow the threshold itself to smoothly vary across the image.

2.4 Binarization

Document binarization is a preprocessing task, considered very important in the term of document analysis systems.(Mitianoudis and Papamarkos, 2015) It automatically converts the document images in a bi-level form in such way that the foreground information is represented by black pixels and the background by white ones.