





PROJECT REPORT  
Hangeul Optical Character Recognition  
on JPEG File

Albert Budi Christian

09.02.0001

2013

	<b>PERPUSTAKAAN</b>
NO. INV : 240 / S / IK / C 1	
TGL : 06 / 02 / 2014	
PARAF : 	

FACULTY OF COMPUTER SCIENCE

SOEGIJAPRANATA CATHOLIC UNIVERSITY

Jl. Pawiyatan Luhur IV/1, Bendan Duwur, SEMARANG 50234

Telp. 024-8441555 (hunting) Web: <http://www.unika.ac.id>

<http://ikomunika.web.id/>

APPROVAL AND RATIFICATION PAGE

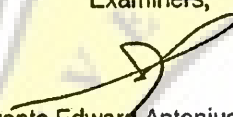
PROJECT REPORT

Hangeul Optical Character Recognition on JPEG File


This project report has been approved and ratified by the Dean of Faculty of Computer Science and Supervisor on January, 21<sup>th</sup> 2014

With approval,

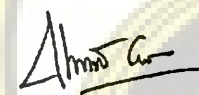
Examiners,

  
Suyanto Edward Antonius, Ir., M.Sc  
NPP : 058.1.1992.116

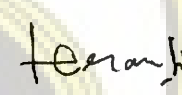
Supervisor,

  
Hironimus Leong, S.Kom., M.Kom  
NPP : 058.1.2007.273

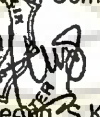
Examiners,

  
Shinta Estri Wahyuningrum, S.Si, M.Cs  
NPP : 058.1.2007.272

Examiners,

  
Rosita Herawati, ST., MIT  
NPP : 058.1.2004.263

Dean of Faculty of Computer Science,

  
Hironimus Leong, S.Kom., M.Kom  
NPP : 058.1.2007.273

## STATEMENT OF ORIGINALITY

I, the undersigned:

Name : Albert Budi Christian

ID : 09.02.0001

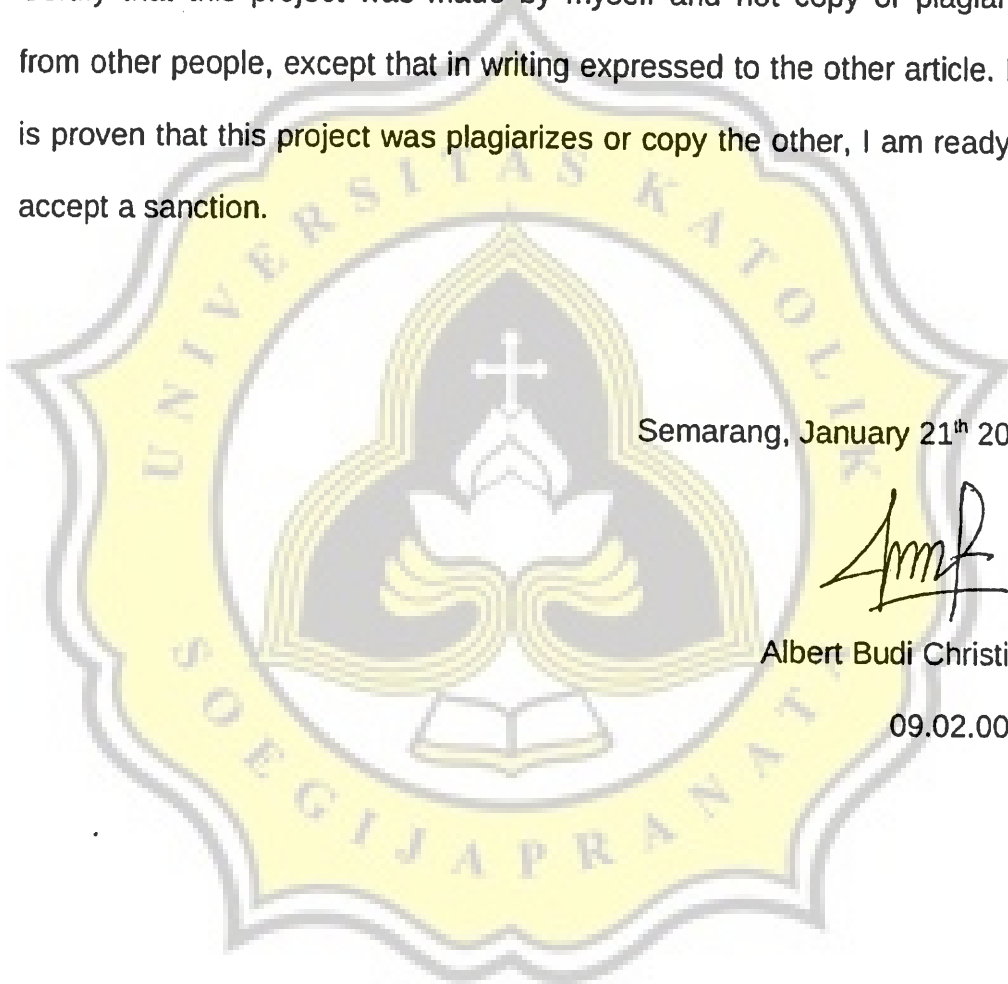
Certify that this project was made by myself and not copy or plagiarize from other people, except that in writing expressed to the other article. If it is proven that this project was plagiarizes or copy the other, I am ready to accept a sanction.

Semarang, January 21<sup>th</sup> 2014



Albert Budi Christian

09.02.0001



## FOREWORD

After four years, I have studied in Faculty of Computer Science of Soegipranata Catholic University. Finally, I have been able to make a Project as a requirement of my graduation. Through this project, I have learned a lot of experience and knowledge of Hangeul Optical Character Recognition on JPEG File.

When I made this project, I have been helped by a lot of people. In this opportunity I would say thank you to:

1. My Lord Jesus Christ, for His blessing and His miracle so I can finish my project.
2. My Parents (PF. Teguh Santoso and FX. Yenny Poerwoko) for their support, love, and pray.
3. My Brother (Andre Budi Christian) who never tired give me motivation everyday.
4. My Girl Friend (Stefanie Karsodihardjo) who always encourage me to finish this project as soon as possible.
5. My Project Supervisor (Hironimus Leong, S. Kom., M. Kom) for his advice and motivation.
6. All of my lectures in Faculty of Computer Science whom teaching and giving a lot of knowledge when I studied in this faculty.
7. "Go HK! Backpacker group" member, my best friend Ferdinand Budi Kurniawan, who relentless give me motivation.
8. My other lectures and friends who support me that can't be mentioned one by one.

## ABSTRACTION

Language is the most relevant way to create and communicate knowledge among human. In order to understand language of each country, it should be started from its alphabet. Korean Alphabet has known as Hangeul. It is one of the unique language that most of young people want to learn. People, who use Latin letter, say that learning the Hangeul is not easy. Therefore, the existence of technology application is needed to solve this problem. It is by creating Optical Character Recognition (OCR). In this project, the writer implement the OCR by using the Java Language Programming.

The method to be applied to this project is Heuristic and Topological Skeleton Algorithm. Here, the Heuristic Algorithm that the writer made is based on the line character, which each Hangeul letter has. The recognition process systems are thinning image of hangeul, detecting the property of line, and giving the result as text.

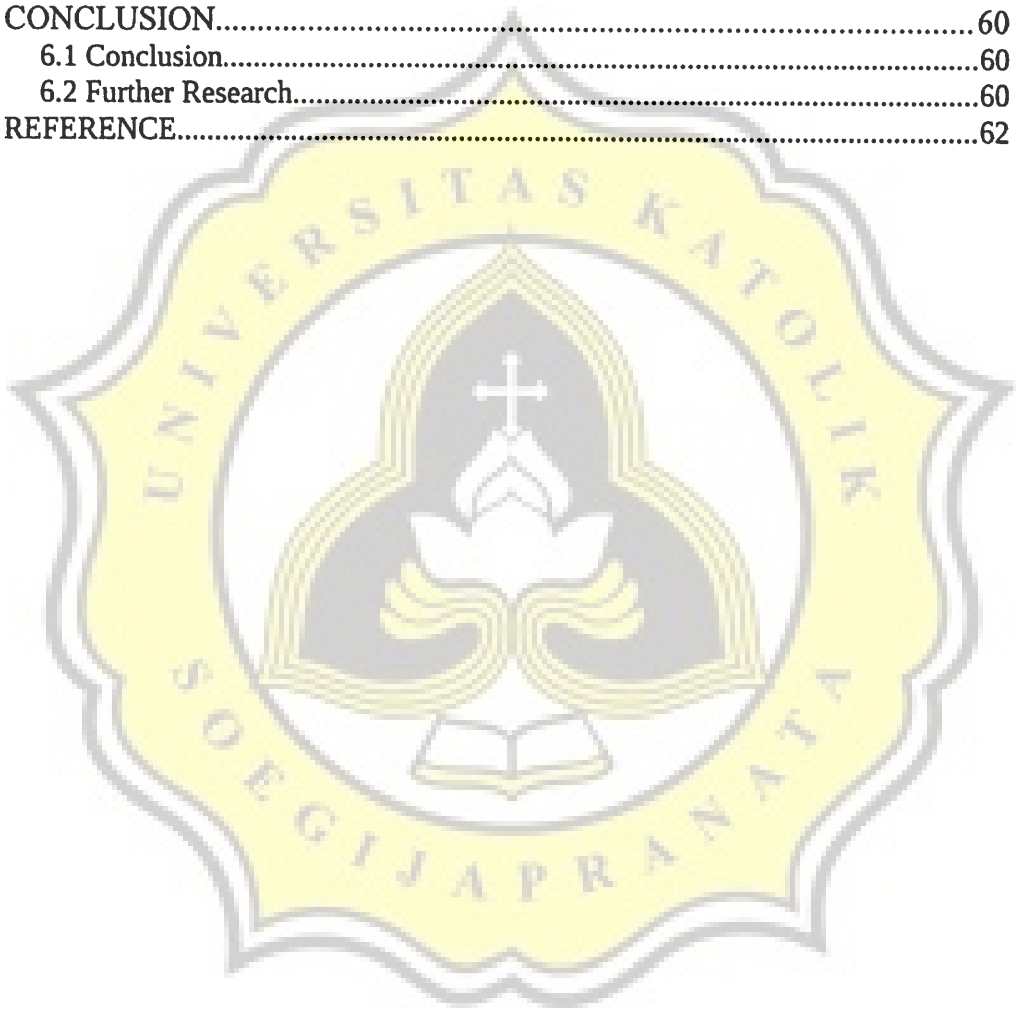
The result of this project is the latin text, which is recognized from the hangeul image that taken from user camera.

**Keyword:** *Hangeul recognition, Heuristic, Topological Skeleton*

## TABLE OF CONTENTS

TITLE.....	i
APPROVAL AND RATIFICATION PAGE.....	ii
STATEMENT OF ORIGINALITY.....	iii
FOREWORD.....	iv
ABSTRACTION.....	v
TABLE OF CONTENTS.....	vi
TABLE OF FIGURE.....	viii
TABLE OF TABLE.....	x
CHAPTER I.....	1
Introduction.....	1
1.1 Background.....	1
1.2 Scope.....	4
1.3 Objective.....	5
CHAPTER II.....	6
Literature Study.....	6
2.1 Data Structure.....	6
2.1.1 Linked List.....	6
2.1 Algorithm.....	9
2.1.1 Cropping.....	9
2.1.2 Topological Skeleton.....	10
2.1.3 Heuristic Algorithm.....	12
CHAPTER III.....	14
Planning.....	14
3.1 Research Methodologies.....	14
3.2 Project Management.....	16
CHAPTER IV.....	17
Analysis and Design.....	17
4.1 Analysis.....	17
4.1.1 Step by step the recognition process.....	17
4.1.2 Use Case Diagram.....	26
4.1.3 Flow Chart.....	27
4.2 Design.....	32
4.2.1 Graphical User Interface(GUI) Design.....	32
4.2.2 Communication Class Diagram.....	33
4.2.3 Class Diagram Detail.....	34
CHAPTER V.....	39
Implementation and Testing.....	39
5.1 Implementation.....	39
5.1.1 Graphical User Interface.....	39

5.1.2 Process Image.....	40
5.2 Testing.....	51
5.2.1 The Main Interface Window.....	52
5.2.2 The Appearance after push Cari Button.....	53
5.2.3 The Appearance after using Cari Button.....	53
5.2.4 The Appearance after using Deteksi Button.....	54
5.2.5 The Experiments Result.....	54
CHAPTER VI.....	60
CONCLUSION.....	60
6.1 Conclusion.....	60
6.2 Further Research.....	60
REFERENCE.....	62





## TABLE OF FIGURE

Figure 2.1 Illustration of Linked List.....	7
Figure 2.2 Example Implementation Topological Skeleton Algorithm....	12
Figure 2.3 Illustration Recognition.....	13
Figure 4.1 Count black pixel illustration.....	18
Figure 4.2 The result of get corner illustration.....	18
Figure 4.3 The result of cropping illustration.....	19
Figure 4.4 Topological skeleton algorithm illustration.....	20
Figure 4.5 Scan black pixel horizontal illustration.....	21
Figure 4.6 Distinguish line horizontal illustration.....	22
Figure 4.7 Scan black pixel vertical illustration.....	22
Figure 4.8 Distinguish line vertical illustration.....	23
Figure 4.9 Choose reference line illustration.....	23
Figure 4.10 The image position coordinate illustration.....	24
Figure 4.11 The important image point coordinate illustration.....	25
Figure 4.12 The determine of position property illustration.....	25
Figure 4.13 Use case diagram.....	26
Figure 4.14 Diagram activity input.....	27
Figure 4.15 Diagram activity black and white.....	27
Figure 4.16 Diagram activity cropping.....	28
Figure 4.17 Diagram activity first step recognition.....	29
Figure 4.18 Diagram activity second step program.....	30
Figure 4.19 Diagram activity final step recognition.....	31
Figure 4.20 Graphical user interface design.....	32
Figure 4.21 Communication class diagram.....	33
Figure 4.22 Utama Class.....	34
Figure 4.23 UploadGambar class.....	35



Figure 4.24 ProcessImage Class.....	36
Figure 4.25 Crop class.....	36
Figure 4.26 Heuristik class.....	37
Figure 4.27 Thinner class.....	37
Figure 4.28 RotateImage class.....	38
Figure 4.29 Point class.....	38
Figure 5.1 Class and constructor GUI.....	40
Figure 5.2 Class searchData.....	41
Figure 5.3 Class prosesData.....	41
Figure 5.4 Method blackWhite.....	42
Figure 5.5 Method cariPixel.....	43
Figure 5.6 Method getPerHuruf.....	44
Figure 5.7 Method thinnerImage.....	45
Figure 5.8 Class Thinner.....	45
Figure 5.9 Method deteksiTitik.....	46
Figure 5.10 Method cekMiring.....	47
Figure 5.11 Method cekKanan.....	47
Figure 5.12 Method cekBawah.....	48
Figure 5.13 Method max.....	48
Figure 5.14 Method analisaHuruf.....	49
Figure 5.15 Method hitungSudut.....	50
Figure 5.16 Method rotasiGambar.....	50
Figure 5.17 The appearance of program.....	52
Figure 5.18 The program appearance after push cari button.....	53
Figure 5.19 The program appearance after using cari button.....	53
Figure 5.20 The program appearance after using deteksi button.....	54

## TABLE OF TABLE

Table 3.1 Project Management .....	16
Table 5.1 Testing Table.....	51
Table 5.2 Result Of Experiments.....	54

