

**STUDI EKSPERIMENTASI KEAKURATAN *TESSERACT-OCR* DALAM
PENGENALAN HURUF KOREA (*HANGEUL*)**

Skripsi sebagai salah satu syarat
memperoleh Derajat Sarjana S1

Program Studi Teknik Informatika



Diajukan oleh :

Nurahmad Miftahudin

NIM : 17106050032

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
YOGYAKARTA**

2023

PENGESAHAN TUGAS AKHIR



KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
FAKULTAS SAINS DAN TEKNOLOGI
Jl. Marsda Adisucipto Telp. (0274) 540971 Fax. (0274) 519739 Yogyakarta 55281

PENGESAHAN TUGAS AKHIR

Nomor : B-2171/Un.02/DST/PP.00.9/08/2023

Tugas Akhir dengan judul : STUDI EKSPERIMENTASI KEAKURATAN TESSERACT-OCR DALAM PENGENALAN HURUF KOREA (HANGEUL)

yang dipersiapkan dan disusun oleh:

Nama : NURAHMAD MIFTAHUDIN
Nomor Induk Mahasiswa : 17106050032
Telah diujikan pada : Senin, 07 Agustus 2023
Nilai ujian Tugas Akhir : A-

dinyatakan telah diterima oleh Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta

TIM UJIAN TUGAS AKHIR



Ketua Sidang

Ir. Maria Ulfah Siregar, S.Kom., MIT., Ph.D.
SIGNED

Valid ID: 64deebff1a73e



Penguji I

Nurochman, S.Kom., M.Kom
SIGNED

Valid ID: 64dab3e1785a



Penguji II

Eko Hadi Gunawan, M.Eng.
SIGNED

Valid ID: 64d1e60b99a2



Yogyakarta, 07 Agustus 2023
UIN Sunan Kalijaga
Dekan Fakultas Sains dan Teknologi

Prof. Dr. Dra. Hj. Khurul Wardati, M.Si.
SIGNED

Valid ID: 64defe6815bac

SURAT PERSETUJUAN SKRIPSI/TUGAS AKHIR



Universitas Islam Negeri Sunan Kalijaga



FM-UINSK-BM-05-03/R0

SURAT PERSETUJUAN SKRIPSI/TUGAS AKHIR

Hal : Persetujuan Skripsi / Tugas Akhir Lamp :

Kepada
Yth. Dekan Fakultas Sains dan Teknologi
UIN Sunan Kalijaga Yogyakarta
di Yogyakarta

Assalamu 'alaikum wr. wb.

Setelah membaca, meneliti, memberikan petunjuk dan mengoreksi serta mengadakan perbaikan seperlunya, maka kami selaku pembimbing berpendapat bahwa skripsi Saudara:

Nama : Nurahmad Miftahudin
NIM : 17106050032
Judul Skripsi : Studi Eksperimentasi Keakuratan *Tesseract-OCR* Dalam
Pengenal Huruf Korea (*Hangeul*)

sudah dapat diajukan kembali kepada Program Studi Teknik Informatika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta sebagai salah satu syarat untuk memperoleh gelar Sarjana Strata Satu dalam Program Studi Teknik Informatika

Dengan ini saya mengharap agar skripsi/tugas akhir tersebut di atas dapat segera dimunaqsyahkan. Atas perhatiannya saya ucapkan terima kasih.

Wassalamu 'alaikum wr. wb.

Yogyakarta, 26 Juli 2023
Pembimbing,

Ir. Maria Ulfah Siregar, S.Kom., M.IT., Ph.D.
NIP. 19780106 200212 2 001

PERNYATAAN KEASLIAN SKRIPSI

SURAT PERNYATAAN KEASLIAN SKRIPSI

Yang bertanda tangan di bawah ini:

Nama : Nurahmad Miftahudin

NIM : 17106050032

Jurusan : Teknik Informatika

Fakultas : Sains dan Teknologi

Menyatakan bahwa:

1. Skripsi yang berjudul: **Studi Eksperimentasi Keakuratan Tesseract-OCR Dalam Pengenalan Huruf Korea (Hangeul)** merupakan hasil karya asli saya yang diajukan untuk memenuhi salah satu persyaratan memperoleh gelar sarjana strata satu (S1) di jurusan Teknik Informatika, Fakultas Sains dan Teknologi, UIN Sunan Kalijaga, Yogyakarta.
2. Semua sumber yang saya gunakan dalam penulisan skripsi ini telah saya cantumkan sesuai dengan ketentuan yang berlaku. Jika di kemudian hari terbukti bahwa karya ini bukan hasil karya saya atau hasil plagiat dari orang lain, maka saya bersedia menerima sanksi yang berlaku.

Yogyakarta, 26 Juli 2023

Yang menyatakan,

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA



Nurahmad Miftahudin
NIM. 17106050032

KATA PENGANTAR

Assalamu'alaikum Wr. Wb.

Alhamdulillahilahirabil'alamin Segala puji bagi Allah SWT yang telah melimpahkan rahmat, hidayah dan karunia-Nya, sehingga penulis dapat menyelesaikan skripsi dengan judul “Studi Eksperimentasi Keakuratan *Tesseract-OCR* Dalam Pengenalan Huruf Korea (*Hangeul*)” dengan lancar. Sholawat serta salam senantiasa tercurahkan kepada Nabi kita Rasulluah SAW, semoga kita semua mendapatkan syafaat di hari kiamat kelak.

Skripsi ini disusun untuk memenuhi sebagian persyaratan mendapatkan gelar kesarjanaan pada Program Studi Teknik Informatika UIN Sunan Kalijaga Yogyakarta. Penulis menyadari bahwa poses penyelesaian tidak lepas dari bantuan dan dukungan dari berbagai pihak baik secara langsung maupun tidak langsung. Oleh karena itu, dengan segala kerendahan hati penulis mengucapkan terima kasih kepada :

1. Bapak Prof. Dr.Phil. Al Makin, S.Ag., M.A., selaku Rektor UIN Sunan Kalijaga Yogyakarta.
2. Ibu Dr. Khurul Wardati, M.Si., selaku Dekan Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta.
3. Ibu Ir. Maria Ulfah Siregar, S.Kom., MIT., Ph.D., selaku Ketua Program Studi Teknik Informatika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta.
4. Ibu Ir. Maria Ulfah Siregar, S.Kom., MIT., Ph.D., selaku Dosen Pembimbing Akademik.
5. Ibu Ir. Maria Ulfah Siregar, S.Kom., MIT., Ph.D.,selaku Dosen Pembimbing Skripsi yang telah membimbing serta memberikan saran dan koreksi kepada penulis selama pengerjaan skripsi ini.
6. Seluruh Dosen Program Studi Teknik Informatika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta yang telah menyalurkan ilmu pengetahuan selama masa perkuliahan.

7. Seluruh Staf Program Studi Teknik Informatika Fakultas Sains dan Teknologi yang telah membantu dalam bidang administrasi.
8. Kedua orang tua dan segenap keluarga yang telah memberikan kasih sayang dan dukungan yang tak terhingga kepada penulis.
9. Seluruh teman-teman Teknik Informatika Angkatan 2017 yang telah menyemangati dan menemani selama masa perkuliahan.
10. Semua pihak yang tidak dapat penulis sebutkan satu persatu yang telah membantu penulis.

Semoga Allah SWT membalas dengan balasan yang terbaik. Penulis menyadari dalam penulisan penelitian ini, terdapat banyak kesalahan, oleh karena itu penulis memohon maaf serta menerima saran dan kritik yang sangat penulis perlukan untuk proses pembelajaran penulis. Semoga penelitian ini dapat bermanfaat bagi pembaca.

Wassalamu'alaikum Warahmatullahi Wabarakatuh

Purworejo, 26 Juni 2023
Penulis

Nurahmad Miftahudin
17106050032

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

HALAMAN PERSEMBAHAN

Alhamdulillahirabbil'alamin. Segala Puji bagi Allah SWT. Terima kasih untuk semua pihak yang telah banyak membantu penulis sampai saat ini. Oleh karena itu penulis ingin mempersembahkan hasil penulisan ini kepada semua orang yang telah banyak membantu, mendukung dan menginspirasi penulis.

Penulis mempersembahkan skripsi ini kepada :

- Orang tua, Bapak Miftakhul Azis dan Heru Yulianto serta Ibu Nur Azizah yang senantiasa memberikan dukungan, kasih sayang dan doanya kepada penulis.
- Kakak, Keponakan dan Saudara yang selalu memberikan semangat.
- Keluarga besar yang turut mendoakan kemudahan bagi penulis.
- Ibu Ir. Maria Ulfah Siregar, S.Kom., MIT., Ph.D., selaku Dosen Pembimbing Skripsi yang telah memberikan banyak pelajaran pada penulis.
- Segenap jajaran dosen Teknik Informatika UIN Sunan Kalijaga, khususnya dosen-dosen pengampu Teknik Informatika angkatan 2017
- Karyawan dan pegawai UIN Sunan Kalijaga yang selalu bersedia membantu penulis ketika terjadi masalah.
- Seluruh teman-teman Teknik Informatika 2017 yang tidak dapat penulis sebutkan satu-persatu.
- Semua pihak yang membantu penulis, baik secara langsung maupun tidak langsung dalam penyusunan skripsi.

MOTTO

“Selama masih bisa melakukannya sendiri lakukanlah sendiri.”



STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

DAFTAR ISI

PENGESAHAN TUGAS AKHIR	ii
SURAT PERSETUJUAN SKRIPSI/TUGAS AKHIR	iii
PERNYATAAN KEASLIAN SKRIPSI	iv
KATA PENGANTAR	v
HALAMAN PERSEMBAHAN	vii
MOTTO	viii
DAFTAR ISI	ix
DAFTAR GAMBAR	xi
DAFTAR TABEL	xii
INTISARI	xiii
ABSTRACT	xiv
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian.....	3
1.6 Kebaruan Penelitian.....	4
BAB II TINJAUAN PUSTAKA DAN LANDASAN TEORI	5
2.1 Tinjauan Pustaka.....	5
2.2 Landasan Teori	9
2.2.1 Bahasa Korea.....	9
2.2.2 Tesseract-OCR	14
2.2.3 Pelatihan Tesseract	14
2.2.4 Pengenalan Pola.....	17
2.2.5 Pengolahan Citra Digital.....	17
2.2.6 Citra Digital.....	19
BAB III METODE PENELITIAN	21
3.1 Studi Pustaka	21

3.2	Pengumpulan Data.....	21
3.3	Metode Pengembangan.....	21
3.4	Alat dan Bahan	23
BAB IV HASIL DAN PEMBAHASAN.....		24
4.1	Image Processing.....	24
4.2	Tahap Pengenalan Huruf.....	28
4.3	Pengujian Data Uji.....	30
BAB V KESIMPULAN DAN SARAN.....		32
5.1	Kesimpulan.....	32
5.2	Saran	32
DAFTAR PUSTAKA.....		34
LAMPIRAN		36



DAFTAR GAMBAR

Gambar 2.1 Daftar Huruf Vokal Korea	11
Gambar 2.2 Daftar Huruf Konsonan Korea	12
Gambar 2.3 Alur Pelatihan Tesseract	16
Gambar 3.1 Diagram Alir Sistem	22
Gambar 4.1 Contoh Data Uji Sebelum Diproses.....	24
Gambar 4.2 Source Code untuk mengubah gambar ke Grayscale Python Open CV	24
Gambar 4.3 Hasil Grayscale	25
Gambar 4.4 Source Code Threshold Python Open CV	25
Gambar 4.5 Hasil Thresholding.....	26
Gambar 4.6 Source Code Gaussian Blur.....	26
Gambar 4.7 Hasil Gaussian Blur	27
Gambar 4.8 Source Code Canny Edge Detection.....	27
Gambar 4.9 Citra Hasil Canny Edge Detection.....	28
Gambar 4.10 Source Code untuk membuka gambar	29
Gambar 4.11 Source Code Config Tesseract-OCR 1	29
Gambar 4.12 Source Code Config Tesseract-OCR 2	29
Gambar 4.13 Source Code untuk Deteksi Huruf.....	29



DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu..... 6



STUDI EKSPERIMENTASI KEAKURATAN *TESSERACT-OCR* DALAM PENGENALAN HURUF KOREA (*HANGEUL*)

Nurahmad Miftahudin

NIM. 17106050032

INTISARI

Banyaknya peminat tentang budaya Korea yang sudah berlangsung lama menarik minat peneliti untuk melakukan penelitian ini, walaupun sudah lama budaya Korea menarik perhatian negara lain salah satunya adalah grup idol Korea. Huruf korea kemudian menarik minat penulis untuk melakukan pengujian pengenalan huruf pada subtitle salah satu *MV* atau *Music Video* grup wanita yaitu *fromis_9* yang berjudul *DM*.

Dalam penelitian ini penulis menggunakan *Tesseract-OCR*. Penulis mencoba menggunakan *Tesseract-OCR* untuk melakukan pengujian pengenalan huruf Korea (*Hangeul*) dalam lirik musik video *fromis_9* yang berjudul *DM*. Peneliti menggunakan *Tesseract-OCR* untuk melihat seberapa akurat *Tesseract-OCR* dapat mengenali huruf korea yang memiliki latar belakang gambar berwarna dan telah dilakukan *image processing*. Data uji yang digunakan berupa 207 tangkapan layar dari *MV fromis_9 DM* yang tercantum subtitle lirik berbahasa Korea dan terdapat juga huruf alfabet dan terbagi menjadi tiga ukuran huruf yaitu 18pt, 24pt, dan 48 yang masing-masing ukuran huruf memiliki 69 gambar uji. Gambar data uji tersebut kemudian juga di proses menggunakan beberapa *image processing* seperti *grayscale*, *thresholding*, dan *gaussian blur*, dan *canny*.

Pada penelitian ini penulis mengelompokkan data uji menjadi tiga, yaitu huruf korea dan alfabet, hanya huruf korea, hanya alfabet kemudian diperoleh hasil dengan persentase keberhasilan dalam melakukan pengenalan huruf dan benar terbaik sebesar 100% dan persentase terburuk 0%. Pada kelompok huruf korea dan alfabet ukuran huruf 24pt memiliki tingkat keberhasilan paling tinggi yaitu 67% dan ukuran huruf 48pt terendah sebesar 53% dan ukuran huruf 18pt berada di tengah dengan persentase 63%. Untuk kelompok hanya korea ukuran huruf 18pt mendapat 65%, ukuran huruf 24pt 66%, dan ukuran huruf 48pt 56%. Kelompok hanya alfabet dianggap berhasil jika tidak ada kotak yang muncul, karena fokus penelitian hanya untuk mendeteksi huruf korea, ada beberapa kasus gambar yang terdeteksi satu kotak pada kata huruf alfabet akan tetapi lebih banyak yang tidak muncul kotak.

Kata Kunci : *Tesseract-OCR*, Huruf Korea, *image processing*

STUDI EKSPERIMENTASI KEAKURATAN *TESSERACT-OCR* DALAM PENGENALAN HURUF KOREA (*HANGEUL*)

Nurahmad miftahudin

NIM. 17106050032

ABSTRACT

Many people interested in Korean culture, which has been going on for a long time, because of this attracted the interest of researchers to conduct this research, even though Korean culture has long attracted the attention of other countries, one of which is the Korean idol group. The Korean letter then attracted the writer's interest in conducting a letter recognition test on the subtitles of one of the women's group MVs or Music Videos, namely fromis_9 entitled DM.

In this study the authors used Tesseract-OCR. The author uses Tesseract-OCR to test the recognition of Korean characters (Hangeul) in the lyrics of the fromis_9 music video entitled DM. Researchers use Tesseract-OCR to see how accurately Tesseract-OCR can identify Korean letters that have a colored image background and have been image processed. The test data used is in the form of 207 screenshots from the MV fromis_9 DM which contains Korean lyric subtitles and there is also an alphabet and is divided into three font sizes, namely 18pt, 24pt and 48, each font size has 69 test images. The test data images were also processed using several image processing such as grayscaling, thresholding, and gaussian blur, and canny.

In this study the authors grouped the test data into a three different category, Korean letters and the alphabet, only Korean letters, and only the alphabet. In the Korean and alphabet letter group, the 24pt font size has the highest success rate, it's gets 67% and the lowest is 48pt font size that gets 53% and the 18pt font size is in the middle with a percentage of 63%. For the Korean only category, the 18pt font size gets 65%, 24pt font size gets 66%, and 48pt font size gets 56%. The alphabet only group was considered successful if no boxes appeared, because the focus of the study was only to detect Korean letters, there were several cases of images that detected one box in the word alphabet but more of them did not appear in the box.

Keyword : *Tesseract-OCR*, Korean font, *image processing*

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Banyaknya peminat tentang budaya Korea yang sudah berlangsung lama menarik minat peneliti untuk melakukan penelitian ini, walaupun sudah lama budaya Korea menarik perhatian negara lain salah satunya adalah grup idol Korea. Grup idol Korea sangat disenangi oleh para remaja Indonesia, menjadikan para remaja tersebut juga mempelajari bahasa korea itu sendiri. Penayangan video-video *K-Pop* di YouTube, Indonesia berada pada posisi ke-2 dengan jumlah persentase 9.9%, sedangkan Korea menempati peringkat pertama dengan persentase 10.1% (Egsaugm 2020). Banyaknya *fans* itu membuat adanya kebutuhan pengenalan huruf, pengenalan huruf menjadi lebih menarik minat jika melalui lagu dari grup *K-Pop*. Selain alasan itu huruf korea kemudian menarik minat penulis untuk melakukan pengujian pengenalan huruf pada subtitle salah satu *MV* atau *Music Video* grup wanita yaitu *fromis_9* yang berjudul *DM*. Huruf Korea sendiri sedikit unik karena cara penulisannya yang seperti ditumpuk dalam satu karakter walau dalam satu karakter itu terdiri dari beberapa huruf.

Dalam penelitian ini penulis menggunakan *Tesseract-OCR*, penulis mencoba menggunakan *Tesseract-OCR* untuk melakukan pengenalan huruf Korea (Hangeul) dalam lirik musik video *fromis_9* yang berjudul *DM*. Peneliti menggunakan *Tesseract-OCR* untuk melihat seberapa akurat *Tesseract-OCR* dapat mengenali huruf korea yang memiliki latar belakang gambar berwarna dan telah dilakukan *image processing*, apakah masih memiliki akurasi tinggi seperti saat melakukan pengenalan huruf dengan latar belakang gambar hitam atau putih. Data uji yang digunakan berupa 207 tangkapan layar dari *MV* *fromis_9* *DM* yang tercantum subtitle lirik berbahasa Korea, 207 gambar tersebut kemudian juga di proses menggunakan beberapa *image processing* seperti *grayscale*, *thresholding*, *gaussian blur*, dan *canny edge detection*.

Tesseract merupakan perangkat lunak *OCR* (*Optical Character Recognition*) yang dikembangkan oleh *HP* pada tahun sekitar 1984-1994 yang

digunakan untuk mengekstrak teks dari gambar. *Tesseract* sendiri merupakan *open source* dan dapat digunakan melalui *command prompt*. Tahapan yang dilakukan penulis berupa penginstalan *Tesseract-OCR*, pembuatan data uji, eksperimen dengan *Tesseract-OCR* untuk komparasi akurasi pengenalan huruf korea.

Dengan menggunakan *Tesseract-OCR* ini diharapkan mendapatkan hasil yang dapat digunakan untuk para peneliti lain dalam hal mendapatkan data pengenalan huruf Korea dengan latar belakang gambar berwarna dan gambar yang telah mendapatkan *image processing*. Pengujian yang dilakukan terbilang cukup rumit karena terdapat huruf alfabet pada beberapa lirik sehingga berpengaruh pada hasil deteksi. Pengujian diutamakan untuk mendeteksi hanya huruf/kata korea saja dengan cara memberi kotak pada huruf/kata yang terdeteksi.

1.2 Rumusan Masalah

Berdasar latar belakang yang telah dijelaskan, rumusan masalah penelitian ini adalah :

1. Berapa rata-rata persentase keberhasilan yang dihasilkan dalam mengenali huruf korea (Hangeul) jika hanya terdapat huruf korea saja, terdapat huruf korea dan alfabet, dan hanya huruf alfabet dalam lirik video musik fromis 9 DM?
2. Bagaimana pengaruh ukuran huruf pada hasil deteksi huruf korea, dan berapa ukuran yang paling optimal dari ukuran huruf 18pt, 24pt, dan 48pt?

1.3 Batasan Masalah

Dalam penelitian ini diperlukan batas masalah sehingga tujuan dari penelitian ini lebih fokus. Adapun batasan masalah yang dibahas dalam penelitian ini adalah:

1. Menggunakan *Tesseract-OCR* v5.0.0.
2. Gambar uji yang digunakan menggunakan format .tif 300dpi.

3. Hanya mendeteksi huruf/kata Korea (Hangeul).
4. Data uji yang diambil berupa tangkapan layar sebanyak 207 gambar dari video musik fromis_9 – DM yang disertai subtitle lirik bahasa Korea (Hangeul).
5. Ukuran huruf subtitle lirik yang dipakai adalah 18pt, 24pt, dan 48pt.
6. Huruf yang dipakai adalah Arial.
7. Penerapan menggunakan bahasa pemrograman Python.

1.4 Tujuan Penelitian

Tujuan penelitian ini adalah :

1. Mengetahui keakuratan *Tesseract-OCR* dalam melakukan pengenalan huruf korea (Hangeul) dalam lirik video musik fromis_9 – DM saat tidak terdapat huruf alfabet dan hanya terdapat huruf korea (Hangeul) saja, terdapat huruf korea dan alfabet, dan hanya alfabet.
2. Mengetahui pengaruh ukuran huruf pada tingkat keakuratan *Tesseract-OCR* dan ukuran huruf yang optimal.

1.5 Manfaat Penelitian

Hasil penelitian ini diharapkan memiliki manfaat penelitian sebagai berikut:

1. Bagi penulis, diharapkan dapat membantu untuk mengaplikasikan *Tesseract-OCR* dalam mengenali huruf Korea (Hangeul).
2. Bagi akademisi, diharapkan bisa membantu dalam hal referensi pengenalan huruf Korea (Hangeul) dan wawasan baru dalam pengenalan huruf menggunakan *Tesseract-OCR*.
3. Bagi peneliti lain, diharapkan dapat memberi referensi untuk para peneliti dalam hal keakuratan *Tesseract-OCR* dalam mengenali huruf Korea (Hangeul) dengan latar belakang gambar berwarna dan dilakukan *image processing* serta dalam hal keakuratan *Tesseract-OCR* dalam mengenali huruf Korea (Hangeul) yang juga terdapat huruf alfabet.

1.6 Kebaruan Penelitian

Penelitian tentang “Pengenalan Huruf Korea Menggunakan *Tesseract-OCR* (Studi Kasus Lirik Lagu *MV fromis_9 DM*)” belum pernah dilakukan sebelumnya, terdapat penelitian yang menggunakan *Tesseract-OCR* di Universitas Islam Negeri Sunan Kalijaga Yogyakarta hanya saja objek yang digunakan berbeda, yaitu bahasa Jepang Hiragana. Penelitian lain juga ada beberapa yang menggunakan bahasa Korea akan tetapi hanya menggunakan tulisan tangan dan berupa huruf satuan tanpa kombinasi.



BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Hasil dari pengujian sistem pengenalan huruf Korea menggunakan *Tesseract-OCR* dapat diambil kesimpulan sebagai berikut:

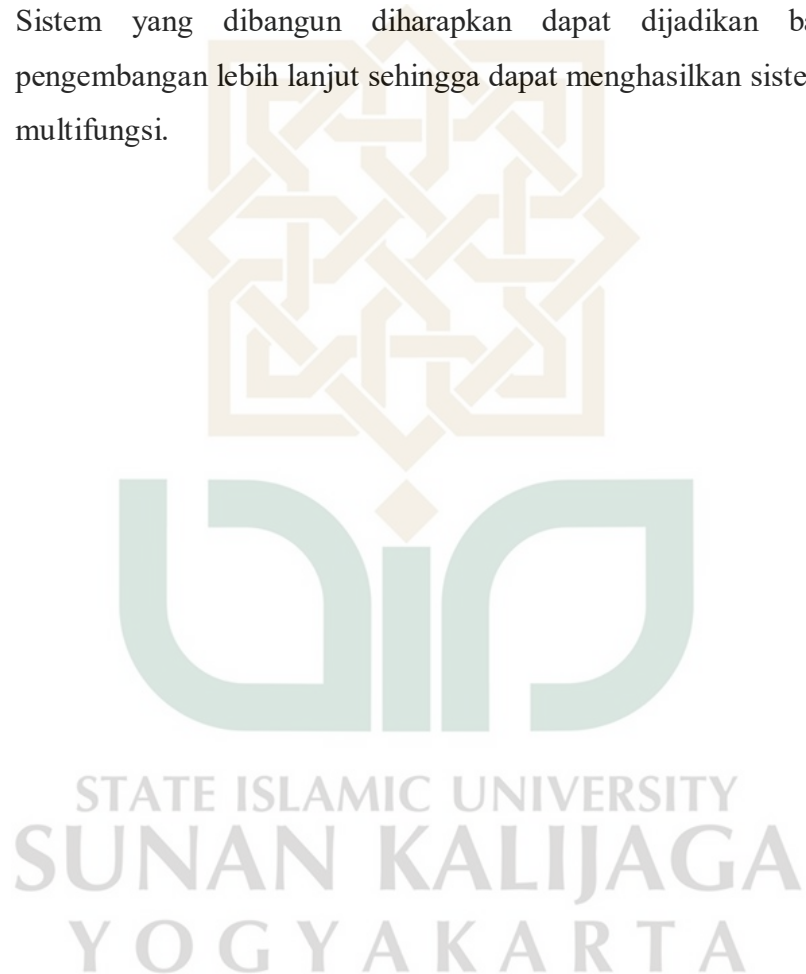
1. Rata-rata persentase keberhasilan yang dihasilkan dalam mengenali huruf korea (Hangeul) jika hanya terdapat huruf korea saja dalam lirik video musik fromis_9 DM sebesar 65% pada ukuran huruf 18pt, 66% ukuran huruf 24pt, dan 56% ukuran huruf 48pt.
2. Rata-rata persentase keberhasilan yang dihasilkan dalam mengenali huruf korea (Hangeul) jika juga terdapat huruf alfabet dalam lirik video musik fromis_9 DM sebesar 63% pada ukuran huruf 18pt, 67% ukuran huruf 24pt, dan 53% ukuran huruf 48pt.
3. Rata-rata keberhasilan yang dihasilkan dalam mengenali huruf korea (Hangeul) jika hanya terdapat huruf alfabet dalam lirik video musik fromis_9 DM sebesar 95% untuk ukuran huruf 18pt, 97% ukuran huruf 24pt, 93% ukuran huruf 48pt, untuk yang tidak ada huruf korea pada data uji dianggap berhasil jika tidak ada kotak yang mendeteksi pada huruf alfabet.
4. Ukuran huruf yang digunakan berpengaruh pada persentase yang dihasilkan.
5. Untuk hasil yang paling optimal didapatkan pada data yang menggunakan ukuran huruf 24pt.

5.2 Saran

Penelitian yang dilakukan tidak lepas dari kekurangan dan kelemahan. Oleh karena itu, pengembangan sistem lebih lanjut perlu diperhatikan beberapa hal diantaranya:

1. Perlunya pengembangan antarmuka agar memudahkan dalam membuka file yang baru, jadi tidak harus menulis nama file secara manual.

2. Masih terdapatnya beberapa data uji yang gagal untuk dibuka, akan lebih baik kalau bisa diminimalisir.
3. Masih munculnya kotak untuk deteksi pada kategori hanya huruf alfabet yang seharusnya tidak muncul sama sekali, akan lebih baik kalau bisa mendapat akurasi 100% yang menandakan tidak munculnya kotak sama sekali.
4. Sistem yang dibangun diharapkan dapat dijadikan bahan dalam pengembangan lebih lanjut sehingga dapat menghasilkan sistem yang lebih multifungsi.



DAFTAR PUSTAKA

- Ali, Abder-Rahman. 2018. "Canny Edge Detector Menggunakan Python." 2018. <https://code.tutsplus.com/id/tutorials/canny-edge-detector-using-python--cms-30095>.
- Egsaugm. 2020. "Fenomena Korean Wave Di Indonesia." 2020. <https://egsa.geo.ugm.ac.id/2020/09/30/fenomena-korean-wave-di-indonesia/>.
- Kumaseh, Max R, Luther Latumakulita, and Nelson Nainggolan. 2013. "Segmentasi Citra Digital Ikan Menggunakan Metode Thresholding." *Jurnal Ilmiah Sains* 13 (1): 74. <https://doi.org/10.35799/jis.13.1.2013.2057>.
- Rostineu. 2014. *30 Hari Belajar Bahasa Korea Dengan Mudah & Lancar*. Edisi Pert. Jakarta: TransMedia Pustaka.
- Sindar, Anita, and R M Sinaga. 2017. "Implementasi Teknik Threshoding Pada Segmentasi Citra Digital." *Jurnal Mantik Penusa* 1 (2): 48–51.
- Smith, Ray. 2007. "An Overview of the Tesseract OCR Engine." *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR 2*: 629–33. <https://doi.org/10.1109/ICDAR.2007.4376991>.
- Soeparno, Haryono, and Triyanto Kun. 2012. "Penerapan Metode Konvolusi (Wikaria Gazali; Dkk) PENERAPAN METODE KONVOLUSI DALAM PENGOLAHAN CITRA DIGITAL." *Mat Stat* 12 (2): 103–13.
- Tabora, Vincent. 2020. "JPEG, TIFF, PNG, SVG File Formats And When To Use Them." 2020. [https://medium.com/hd-pro/jpeg-tiff-png-svg-file-formats-and-when-to-use-them-1b2cde4074d3#:~:text=The PNG \(Portable Network Graphics,is ideal for complex images.&text=PNG.,much quality in the image](https://medium.com/hd-pro/jpeg-tiff-png-svg-file-formats-and-when-to-use-them-1b2cde4074d3#:~:text=The PNG (Portable Network Graphics,is ideal for complex images.&text=PNG.,much quality in the image).
- Wang, Patrick S P. 1993. *Handbook Of Pattern Recognition And Computer Vision*. World Scientific Publishing.

Wedianto, Andre, Herlina Latipa Sari, and Yanolanda Suzantri H. 2016. "269-
Article Text-766-1-10-20160609" 12 (1): 21–30.

