

## DAFTAR ISI

HALAMAN PENGESAHAN PEMBIMBING .....	iii
HALAMAN PENGESAHAN PENGUJI.....	iv
SURAT PERNYATAAN KARYA ASLI TUGAS AKHIR.....	v
PERNYATAAN BEBAS PLAGIASI.....	vi
ABSTRAK .....	vii
<i>ABSTRACT</i> .....	viii
KATA PENGANTAR.....	ix
DAFTAR ISI .....	x
DAFTAR TABEL .....	xii
DAFTAR GAMBAR.....	xiii
DAFTAR MODUL PROGRAM.....	xiv
DAFTAR LAMPIRAN .....	xv
BAB I PENDAHULUAN .....	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah .....	3
1.3. Batasan Masalah.....	3
1.4. Tujuan.....	3
1.5. Manfaat.....	3
1.6. Tahapan Penelitian .....	4
1.7. Sistematika Penulisan.....	5
BAB II TINJAUAN LITERATUR .....	7
2.1. Kartu Tanda Mahasiswa (KTM) .....	7
2.2. <i>Computer Vision</i> .....	7
2.3. Citra Digital .....	8
2.4. <i>Optical Character Recognition (OCR)</i> .....	8
2.4.1. <i>Preprocessing</i> .....	9
2.4.2. Segmentasi.....	13
2.4.3. Ekstraksi Ciri .....	14
2.4.4. <i>Recognition</i> .....	14
2.5. <i>Tesseract OCR</i> .....	16
2.5.1. <i>Adaptive Thresholding and Page Layout Analysis</i> .....	17
2.5.2. <i>Line Finding and Baseline Fitting</i> .....	17
2.5.3. <i>Word Detection and Segmentation</i> .....	18
2.5.4. <i>Character Recognition</i> .....	18
2.6. Evaluasi .....	19
2.7. Penelitian Terdahulu.....	20
BAB III METODOLOGI PENELITIAN DAN PENGEMBANGAN SISTEM .....	24
3.1. Studi Literatur.....	24
3.2. Implementasi Metodologi Penelitian.....	24
3.2.1. Pengumpulan Data.....	25
3.2.2. Preprocessing Data .....	27
3.2.3. Pengenalan Karakter pada KTM menggunakan Tesseract.....	31
3.2.4. Identifikasi NIM dari Hasil Pengenalan Tesseract .....	38
3.2.5. Analisis dan Evaluasi Hasil Pengenalan Tesseract.....	42
3.3. Pengembangan Sistem.....	44
3.3.1. Analisis Kebutuhan Sistem.....	45

3.3.2. Perancangan Desain Sistem.....	45
3.3.3. Perancangan Pengujian Sistem.....	50
BAB IV HASIL DAN PEMBAHASAN.....	52
4.1. Hasil Penelitian.....	52
4.1.1. Pengumpulan Data.....	52
4.1.2. <i>Preprocessing</i> Data .....	53
4.1.3. Pengenalan Karakter pada KTM Menggunakan Tesseract .....	55
4.1.4. Identifikasi NIM dari Hasil Pengenalan Tesseract .....	56
4.1.5. Implementasi Pengembangan Sistem .....	58
4.2. Pengujian .....	59
4.2.1. Pengujian Akurasi .....	59
4.2.2. Pengujian Kecepatan .....	62
4.2.3. Pengujian Sistem .....	63
4.3. Pembahasan .....	63
BAB V KESIMPULAN DAN SARAN .....	66
5.1. Kesimpulan.....	66
5.2. Saran .....	66
DAFTAR PUSTAKA.....	67
LAMPIRAN .....	70

## DAFTAR TABEL

Tabel 2.1. <i>Confusion Matrix</i> .....	19
Tabel 2.2. <i>State of The Art</i> pengenalan karakter .....	20
Tabel 2.3. <i>State of The Art</i> pengenalan karakter lanjutan.....	21
Tabel 2.4. <i>State of The Art preprocessing</i> .....	22
Tabel 2.5. <i>State of The Art preprocessing</i> lanjutan .....	23
Tabel 3.1. Dataframe hasil pengenalan <i>Tesseract-OCR</i> .....	38
Tabel 3.2. Dataframe hasil pembersihan data pengenalan <i>Tesseract-OCR</i> .....	40
Tabel 3.3. Rancangan pengujian kemampuan pengenalan oleh <i>Tesseract-OCR</i> .....	42
Tabel 3.4. Rancangan pengujian pengaruh <i>brightness</i> dan <i>contrast</i> terhadap pengenalan .	43
Tabel 3.5. Rancangan pengujian pengaruh <i>angle correction</i> terhadap pengenalan .....	43
Tabel 3.6. Rancangan pengujian kecepatan pengenalan dan identifikasi sistem .....	43
Tabel 3.7. Rincian tabel data mahasiswa pada basis data .....	49
Tabel 3.8. Skenario pengujian fungsi sistem.....	51
Tabel 4.1. Hasil pengujian pengaruh <i>brightness</i> dan <i>contrast</i> terhadap pengenalan .....	60
Tabel 4.2. Hasil pengujian kemampuan pengenalan oleh <i>Tesseract-OCR</i> .....	61
Tabel 4.3. Hasil pengujian pengaruh <i>angle correction</i> terhadap pengenalan .....	62
Tabel 4.4. Hasil pengujian kecepatan sistem.....	62
Tabel 4.5. Hasil pengujian fungsi sistem.....	63

## DAFTAR GAMBAR

Gambar 2.1. Digitalisasi Citra Kontinu .....	8
Gambar 2.2. Proses OCR Secara Umum .....	8
Gambar 2.3. Ilustrasi parameter $r, \theta$ .....	11
Gambar 2.4. Ilustrasi perpotongan kurva $r, \theta$ .....	11
Gambar 2.5. Arsitektur LSTM .....	15
Gambar 2.6. Arsitektur <i>Tesseract OCR</i> .....	17
Gambar 2.7. Kandidat <i>chop points</i> pada karakter yang terhubung .....	18
Gambar 2.8. Karakter dengan bagian yang terputus atau hilang .....	18
Gambar 3.1. Tahapan penelitian .....	25
Gambar 3.2. Contoh hasil augmentasi data .....	27
Gambar 3.3. Perubahan kecerahan citra dengan nilai (a) 0, (b) -8, (c) -16 .....	28
Gambar 3.4. Perubahan kontras citra dengan nilai (a) 0, (b) +8, (c) +16 .....	28
Gambar 3.5. <i>Flowchart hough transform</i> .....	29
Gambar 3.6. Ilustrasi transformasi hough .....	30
Gambar 3.7. Hasil perubahan citra (a) <i>grayscale</i> , (b) biner .....	31
Gambar 3.8. <i>Flowchart Adaptive Thresholding and Page Layout Analysis</i> .....	33
Gambar 3.9. <i>Flowchart Line Finding and Baseline Fitting</i> .....	34
Gambar 3.10. <i>Flowchart Word Detection and Segmentation</i> .....	35
Gambar 3.11. <i>Flowchart Character Classifier</i> .....	37
Gambar 3.12. <i>Flowchart</i> (a) Identifikasi NIM, (b) <i>Clean data raw</i> .....	39
Gambar 3.13. Ilustrasi <i>key</i> dan <i>value</i> .....	40
Gambar 3.14. Ilustrasi toleransi eror .....	40
Gambar 3.15. Alur pengembangan sistem .....	44
Gambar 3.16. Arsitektur Sistem .....	46
Gambar 3.17. <i>Flowchart</i> sistem .....	47
Gambar 3.18. DFD level 0 .....	47
Gambar 3.19. DFD level 1 .....	48
Gambar 3.20. Rancangan <i>interface</i> halaman <i>input</i> .....	49
Gambar 3.21. Rancangan <i>interface</i> halaman hasil .....	50
Gambar 4.1. Sampel hasil <i>brightness and contrast adjustment</i> .....	53
Gambar 4.2. Sampel hasil <i>angle correction</i> .....	55
Gambar 4.3. Sampel hasil pengenalan <i>Tesseract OCR</i> .....	56
Gambar 4.4. Sampel hasil <i>clean data</i> pengenalan .....	57
Gambar 4.5. Sampel hasil identifikasi <i>value</i> .....	58
Gambar 4.6. Tampilan sistem halaman <i>input</i> .....	58
Gambar 4.7. Tampilan sistem halaman hasil .....	59
Gambar 4.8. Sampel hasil <i>brightness</i> dan <i>contrast</i> .....	60
Gambar 4.9. Sampel hasil perbaikan kemiringan citra .....	61

## DAFTAR MODUL PROGRAM

Modul Program 2.1. Algoritma <i>hough transform</i> .....	12
Modul Program 2.2. Algoritma <i>preprocessing</i> .....	13
Modul Program 2.3. Algoritma LSTM.....	16
Modul Program 4.1. Algoritma augmentasi data .....	52
Modul Program 4.3. Algoritma <i>brightness and contrast adjustment</i> .....	53
Modul Program 4.2. Algoritma <i>angle correction</i> .....	54
Modul Program 4.4. Algoritma pengenalan menggunakan <i>Tesseract OCR</i> .....	55
Modul Program 4.5. Algoritma <i>clean Tesseract OCR data result</i> .....	56
Modul Program 4.6. Algoritma pencarian teks <i>key</i> dan <i>value</i> .....	58

## DAFTAR LAMPIRAN

Lampiran 1. Survey kecepatan pengetikan NIM pegawai administrasi UPNVY .....	71
Lampiran 2. Sampel data citra KTM .....	72
Lampiran 3. Contoh hasil pengenalan oleh <i>Tesseract-OCR</i> .....	73
Lampiran 4. Contoh hasil pembersihan data pengenalan <i>Tesseract-OCR</i> .....	76
Lampiran 5. Contoh hasil identifikasi <i>value</i> berdasarkan nilai tengah <i>key</i> .....	78
Lampiran 6. Hasil pengaruh <i>brightness</i> dan <i>contrast</i> terhadap pengenalan skenario 1 .....	79
Lampiran 7. Hasil pengaruh <i>brightness</i> dan <i>contrast</i> terhadap pengenalan skenario 2 .....	79
Lampiran 8. Hasil pengaruh <i>brightness</i> dan <i>contrast</i> terhadap pengenalan skenario 3 .....	80
Lampiran 9. Hasil pengaruh <i>brightness</i> dan <i>contrast</i> terhadap pengenalan skenario 4 .....	81
Lampiran 10. Hasil pengaruh <i>brightness</i> dan <i>contrast</i> terhadap pengenalan skenario 5 ....	81
Lampiran 11. Hasil pengaruh <i>brightness</i> dan <i>contrast</i> terhadap pengenalan skenario 6 ....	82
Lampiran 12. Hasil pengaruh <i>brightness</i> dan <i>contrast</i> terhadap pengenalan skenario 7 ....	83
Lampiran 13. Hasil pengaruh <i>brightness</i> dan <i>contrast</i> terhadap pengenalan skenario 8 ....	83
Lampiran 14. Hasil pengaruh <i>brightness</i> dan <i>contrast</i> terhadap pengenalan skenario 9 ....	84
Lampiran 15. Hasil pengujian kemampuan pengenalan untuk setiap data KTM.....	85
Lampiran 16. Hasil pengujian pengaruh penggunaan <i>angle correction</i> terhadap identifikasi dan pengenalan karakter <i>value</i> .....	86
Lampiran 17. Hasil pengujian kecepatan pengenalan dan identifikasi sistem .....	88