

PERBANDINGAN HASIL SEGMENTASI KATA TULISAN
TANGAN ARAB MENGGUNAKAN EKSTRAKSI FITUR
ZONING DAN WUMI

*Diajukan Sebagai Syarat untuk Menyelesaikan
Pendidikan Program Strata-1 pada
Jurusan Teknik Informatika*



Oleh :

Achmad Dwi Suriyanto
NIM : 09021181419028

**Jurusan Teknik Informatika
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2018**

LEMBAR PENGESAHAN SKRIPSI

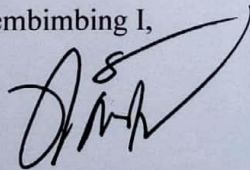
PERBANDINGAN HASIL SEGMENTASI KATA TULISAN
TANGAN ARAB MENGGUNAKAN EKSTRAKSI FITUR
ZONING DAN WUMI

Oleh :

Achmad Dwi Suriyanto
NIM : 09021181419028

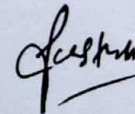
Palembang, Desember 2018

Pembimbing I,



Samsuryadi, M.Kom., Ph.D
NIP 197102041997021003

Pembimbing II,



Desty Rodiah, M.T
NIP 1671016112890005

Mengetahui,
Ketua Jurusan Teknik Informatika



Rifkie Primartha, MT
NIP. 197106012009121004

TANDA LULUS SIDANG TUGAS AKHIR

Pada hari Sabtu, 15 Desember 2018 telah dilaksanakan ujian sidang tugas akhir oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

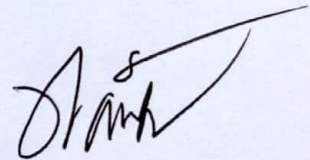
Nama : Achmad Dwi Suriyanto

NIM : 09021181419028

Judul : Perbandingan Hasil Segmentasi Kata Tulisan Tangan Arab Menggunakan Ekstraksi Fitur Zoning dan Wumi

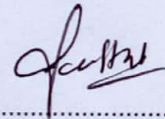
1. Pembimbing I

Samsuryadi, M.Kom., Ph.D
NIP. 197102041997021003



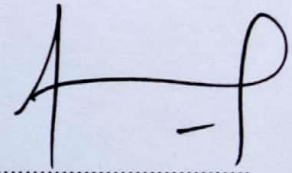
2. Pembimbing II

Desty Rodiah, M.T
NIP. 1671016112890005



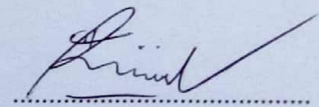
3. Penguji I

M. Fachrurrozi, M.T
NIP. 198005222008121002



4. Penguji II

Mastura Diana Marieska, M.T
NIP. 198603212018032001



Mengetahui,
Ketua Jurusan Teknik Informatika



Rifkie Primartha, MT
NIP. 197706012009121004

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Achmad Dwi Suriyanto

NIM : 09021181419028

Program Studi : Teknik Informatika

Judul Skripsi : Perbaikan Hasil Segmentasi Kata Tulisan Tangan Arab dengan *Random Forests* Menggunakan Ekstraksi Fitur *Zoning* dan WUMI

Hasil Pengecekan *Software iThenticate/turnitin* : 7%

Menyatakan bahwa laporan proyek saya merupakan hasil kerja sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan proyek ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.



Indralaya, 7 Desember 2018



Achmad Dwi Suriyanto
NIM. 09021181419028

MOTTO

- “ Tidak ada daya dan upaya melainkan karena الله ”
- “ Tidaklah seseorang dibebankan melainkan sesuai dengan kesanggupannya ”
- *Faith in deen requires proof, syukr/shabr or kufr*
- Cobalah terus, jika belum bisa, coba lagi

Skripsi ini kupersembahkan kepada :

- الله dan Rasulullah ﷺ
- Kedua Orang Tuaku
- Ummat

ABSTRACT

Word segmentation is one of stages on the handwriting Arabic character recognition process. Characteristics of writing Arabic words and characters are have many variations and similarities between characters, thus the accuracy of segmentation system results is not yet qualified. In this study, comparison of handwritten Arabic words using zoning and WUMI features extraction was compared. The segmentation process applies the method of over segmentation and segmentation validation. The results of handwritten Arabic word segmentation comparison using the zoning and WUMI features extraction, is that the AZZC (Adaptive Zoning Zone Characteristics-based) feature has the highest word segmentation accuracy which is 54.16%.

Keywords : Arabic words segmentation, Over segmentation, Zoning Features, Moment Features, Random Forests, Weighted United Moment Invariant (WUMI), Adaptive Zoning, Static Zoning, Pixel Density, Line Segments, Zone Characteristics.

ABSTRAK

Segmentasi kata merupakan salah satu tahapan dari proses pengenalan karakter Arab tulisan tangan. Karakteristik penulisan kata maupun karakter Arab yang memiliki banyak variasi serta memiliki kemiripan antar karakter, mengakibatkan akurasi dari sistem segmentasi yang dihasilkan belum mumpuni. Pada penelitian ini dilakukan perbandingan hasil segmentasi kata Arab tulisan tangan menggunakan fitur *zoning* dan WUMI. Proses segmentasi menerapkan metode *over segmentation* dan validasi segmentasi. Hasil perbandingan segmentasi kata Arab tulisan tangan menggunakan fitur *zoning* dan WUMI, ialah fitur AZZC (*Adaptive Zoning Zone Characteristics-based*) memiliki tingkat akurasi segmentasi kata tertinggi yaitu 54,16%.

Kata Kunci : Segmentasi kata Arab, *Over segmentation*, Fitur *zoning*, Fitur *moment*, *Random Forests*, *Weighted United Moment Invariant* (WUMI), *Adaptive Zoning*, *Static Zoning*, *Pixel Density*, *Line Segments*, *Zone Characteristics*.

KATA PENGANTAR

Segala puji bagi Allah atas berkah dan rahmat-Nya penulis dapat menyelesaikan skripsi ini. Skripsi ini disusun untuk memenuhi salah satu syarat dalam menyelesaikan pendidikan program Strata-1 pada Fakultas Ilmu Komputer Program Studi Teknik Informatika di Universitas Sriwijaya.

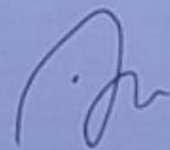
Pada kesempatan ini, penulis ingin menyampaikan ucapan terima kasih kepada pihak-pihak yang telah memberikan dukungan, bimbingan, dan motivasi selama proses penelitian ini dilaksanakan. Secara khusus penulis ingin menyampaikan rasa terima kasih kepada :

1. Kedua orang tuaku, bapak (Sugiharto) dan ibuk (Nanik Lestari) yang telah menuntun, mendoakan, serta memberikan dukungan.
2. Mbakku (Annisa Puji Astuti/tukul) dan adikku (Rahmadi Satrio Wibowo/bedul) yang telah memberikan semangat, doa, dan nasihat.
3. Bapak Jaidan Jauhari, S.Pd., M.T. selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
4. Bapak Rifkie Primartha, S.T., M.T. selaku Ketua Jurusan Teknik Informatika.
5. Bapak Samsuryadi, M.Kom., Ph.D. selaku dosen pembimbing 1 yang dengan sabar membimbing dan menyediakan waktu untuk mengarahkan penulis sehingga skripsi ini dapat diselesaikan.
6. Ibu Desty Rodiah, M.T. selaku dosen pembimbing 2 yang telah memberikan bimbingan dengan sabar, nasihat, motivasi serta telah menyediakan waktu sehingga penulis dapat menyelesaikan skripsi ini.

7. Ibu Yunita, S.SI., M.CS. selaku pembimbing akademik yang telah memberikan bimbingan dalam hal perkuliahan.
8. Seluruh dosen Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya yang telah membagikan ilmunya kepada penulis selama proses perkuliahan.
9. Seluruh staf administrasi dan pegawai yang telah membantu penulis dalam hal administrasi selama proses perkuliahan.
10. Rekan-rekan seperjuangan IF Reg 14 yang tidak dapat disebutkan satu persatu.
11. Rekan-rekan Kost Al-Kudus yang telah membersamai revisian serta semua rekan yang telah memberikan doa dan dukungannya.
12. Semua pihak yang telah membantu dalam penyelesaian skripsi ini.

Penulis menyadari skripsi ini masih jauh dari sempurna. Oleh karena itu, kritik dan saran yang bersifat membangun sangat penulis harapkan. Akhir kata, penulis berharap agar skripsi ini dapat bermanfaat bagi kita semua, insya Allah.

Indralaya, Desember 2018



Penulis

DAFTAR ISI

	Halaman
HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN.....	ii
HALAMAN PERSETUJUAN KOMISI PENGUJI	iii
HALAMAN PERNYATAAN BEBAS PLAGIAT	iv
HALAMAN MOTTO DAN PERSEMBAHAN.....	v
ABSTRACT.....	vi
ABSTRAK	vii
KATA PENGANTAR	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR	xv
DAFTAR ALGORITMA.....	xviii
DAFTAR ISTILAH, SINGKATAN, DAN LAMBANG	xix
BAB I PENDAHULUAN	
1.1 Pendahuluan.....	I-1
1.2 Latar Belakang.....	I-1
1.3 Rumusan Masalah.....	I-4
1.4 Tujuan Penelitian	I-4
1.5 Manfaat Penelitian	I-5
1.6 Batasan Masalah	I-5
1.7 Sistematika Penulisan	I-5
1.8 Kesimpulan	I-6
BAB II KAJIAN LITERATUR	
2.1 Pendahuluan.....	II-1
2.2 Landasan Teori	II-1
2.2.1 Pengenalan Karakter.....	II-1
2.2.2 Segmentasi Karakter.....	II-4
2.2.3 <i>Over Segmentation</i> dan Validasi Segmentasi.....	II-5

2.2.4 Ekstraksi Fitur	II-17
2.3 Penelitian Lain Yang Relevan	II-27
2.3.1 Segmentasi Kata Tulisan Tangan dalam Aksara Arab	II-27
2.3.2 Ekstraksi Fitur dalam Pengenalan Karakter	II-29
2.4 Kesimpulan	II-29

BAB III METODOLOGI PENELITIAN

3.1 Pendahuluan.....	III-1
3.2 Unit Penelitian	III-1
3.3 Pengumpulan Data	III-1
3.3.1 Jenis Data	III-1
3.3.2 Sumber Data	III-2
3.3.3 Metode Pengumpulan Data	III-2
3.4 Tahapan Penelitian.....	III-2
3.4.1 Menentukan Kerangka Kerja / <i>Framework</i>	III-3
3.4.2 Menetapkan Kriteria Pengujian.....	III-5
3.4.3 Menetapkan Format Data Pengujian	III-6
3.4.4 Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian	III-7
3.4.5 Melakukan Pengujian Penelitian	III-7
3.4.6 Melakukan Analisa Hasil Pengujian dan Membuat Kesimpulan.....	III-8
3.5 Metode Pengembangan Perangkat Lunak.....	III-9
3.5.1. <i>Rational Unified Process</i>	III-9
3.5.2. Fase Insepsi	III-10
3.5.3. Fase Elaborasi.....	III-10
3.5.4. Fase Konstruksi	III-10
3.5.5. Fase Transisi	III-10
3.6 Manajemen Proyek Penelitian	III-11

BAB IV PENGEMBANGAN PERANGKAT LUNAK

4.1 Pendahuluan.....	IV-1
4.2 Fase Insepsi.....	IV-1
4.2.1 Pemodelan Bisnis	IV-1
4.2.2 Kebutuhan	IV-5

4.2.3 Analisis dan Desain	IV-7
4.3 Fase Elaborasi	IV-55
4.3.1 Pemodelan Bisnis	IV-55
4.3.2 Kebutuhan	IV-61
4.3.3 Analisis dan Desain	IV-61
4.4 Fase Konstruksi.....	IV-65
4.4.1. Pemodelan Bisnis	IV-65
4.4.2. Kebutuhan	IV-65
4.4.3. Analisis dan Desain	IV-66
4.4.4. Implementasi	IV-81
4.5 Fase Transisi	IV-88
4.5.1 Pemodelan Bisnis	IV-88
4.5.2 Kebutuhan	IV-88
4.5.3 Analisis dan Desain	IV-88
4.5.4 Implementasi	IV-91

BAB V HASIL DAN ANALISA PENELITIAN

5.1 Pendahuluan.....	V-1
5.2 Hasil Percobaan Penelitian	V-1
5.3 Analisa Penelitian	V-15

BAB VI KESIMPULAN DAN SARAN

6.1 Pendahuluan.....	VI-1
6.2 Kesimpulan	VI-1
6.3 Saran	VI-2

DAFTAR PUSTAKA

LAMPIRAN

DAFTAR TABEL

	Halaman
III-1. Tabel Rancangan Hasil Pengujian Sistem Segmentasi Kata	III-6
III-2. Tabel Penjadwalan Penelitian dalam Bentuk <i>Work Breakdown Structure</i> (WBS)	III-12
IV-1. Tabel Kebutuhan Fungsional	IV-8
IV-2. Tabel Kebutuhan Non Fungsional	IV-8
IV-3. Tabel Contoh Data Latih	IV-9
IV-4. Tabel Contoh Data Uji	IV-9
IV-5. Tabel Definisi Aktor	IV-27
IV-6. Tabel Definisi <i>Use Case</i>	IV-27
IV-7. Tabel Skenario <i>Use Case</i> Mengelola Data Pelatihan	IV-29
IV-8. Tabel Skenario <i>Use Case</i> Mengelola <i>Dataset</i> Latih	IV-30
IV-9. Tabel Skenario <i>Use Case</i> Menambah Data Karakter	IV-31
IV-10. Tabel Skenario <i>Use Case</i> Menambah Data SA	IV-33
IV-11. Tabel Skenario <i>Use Case</i> Menghapus Data	IV-34
IV-12. Tabel Skenario <i>Use Case</i> Mengelola Data Pengujian	IV-35
IV-13. Tabel Skenario <i>Use Case</i> Mengelola <i>Dataset</i> Uji	IV-35
IV-14. Tabel Skenario <i>Use Case</i> Mengelola Data Uji	IV-37
IV-15. Tabel Skenario <i>Use Case</i> Melakukan Segmentasi Kata Manual	IV-38
IV-16. Tabel Skenario <i>Use Case</i> Melakukan Pelatihan <i>Random Forests</i>	IV-39
IV-17. Tabel Skenario <i>Use Case</i> Melakukan Segmentasi Kata	IV-41
IV-18. Tabel Skenario <i>Use Case</i> Melakukan Komparasi Hasil Segmentasi ..	IV-42
IV-19. Tabel Skenario <i>Use Case</i> Melihat Hasil Komparasi Hasil Segmentasi	IV-43
IV-20. Tabel Skenario <i>Use Case</i> Melihat Tingkat Akurasi Hasil Segmentasi	IV-44
IV-21. Tabel Implementasi Kelas	IV-81
IV-22. Tabel Rencana Pengujian BB Mengelola Data Pelatihan	IV-88
IV-23. Tabel Rencana Pengujian BB Mengelola <i>Dataset</i> Latih	IV-89
IV-24. Tabel Rencana Pengujian BB Menambah Data Karakter	IV-89
IV-25. Tabel Rencana Pengujian BB Menambah Data SA	IV-89
IV-26. Tabel Rencana Pengujian BB Menghapus Data	IV-89
IV-27. Tabel Rencana Pengujian BB Mengelola Data Pengujian	IV-89
IV-28. Tabel Rencana Pengujian BB Mengelola <i>Dataset</i> Uji	IV-90
IV-29. Tabel Rencana Pengujian BB Mengelola Data Uji	IV-90
IV-30. Tabel Rencana Pengujian BB Melakukan Segmentasi Kata Manual ..	IV-90
IV-31. Tabel Rencana Pengujian BB Melakukan Pelatihan <i>Random Forests</i>	IV-90
IV-32. Tabel Rencana Pengujian BB Melakukan Segmentasi Kata	IV-90
IV-33. Tabel Rencana Pengujian BB Melakukan Komparasi Hasil Segmentasi .	IV-91
IV-34. Tabel Rencana Pengujian BB Melihat Hasil Komparasi Hasil Segmentasi ..	IV-91
IV-35. Tabel Rencana Pengujian BB Melihat Tingkat Akurasi Hasil Segmentasi ..	IV-91
IV-36. Tabel Pengujian BB Mengelola Data Pelatihan	IV-92
IV-37. Tabel Pengujian BB Mengelola <i>Dataset</i> Latih	IV-92

IV-38. Tabel Pengujian BB Menambah Data Karakter	IV-93
IV-39. Tabel Pengujian BB Menambah Data SA	IV-94
IV-40. Tabel Pengujian BB Menghapus Data	IV-94
IV-41. Tabel Pengujian BB Mengelola Data Pengujian	IV-95
IV-42. Tabel Pengujian BB Mengelola <i>Dataset</i> Uji	IV-95
IV-43. Tabel Pengujian BB Mengelola Data Uji	IV-96
IV-44. Tabel Pengujian BB Melakukan Segmentasi Kata Manual	IV-97
IV-45. Tabel Pengujian BB Melakukan Pelatihan <i>Random Forests</i>	IV-98
IV-46. Tabel Pengujian BB Melakukan Segmentasi Kata	IV-99
IV-47. Tabel Pengujian BB Melakukan Komparasi Hasil Segmentasi	IV-100
IV-48. Tabel Pengujian BB Melihat Hasil Komparasi Hasil Segmentasi	IV-100
IV-49. Tabel Pengujian BB Melihat Tingkat Akurasi Hasil Segmentasi	IV-100
V-1. Tabel Hasil Percobaan Pencarian Parameter <i>Static Zoning</i> pada Data Karakter	V-2
V-2. Tabel Hasil Percobaan Pencarian Parameter <i>Adaptive Zoning</i> pada Data Karakter	V-4
V-3. Tabel Hasil Percobaan Pencarian Parameter <i>Static Zoning</i> pada Data SA	V-5
V-4. Tabel Hasil Percobaan Pencarian Parameter <i>Adaptive Zoning</i> pada Data SA	V-7
V-5. Tabel Hasil Ekstraksi Fitur dengan SZPD	V-7
V-6. Tabel Hasil Ekstraksi Fitur dengan SZZC	V-8
V-7. Tabel Hasil Ekstraksi Fitur dengan SZLS	V-8
V-8. Tabel Hasil Ekstraksi Fitur dengan AZPD	V-9
V-9. Tabel Hasil Ekstraksi Fitur dengan AZZC	V-10
V-10. Tabel Hasil Ekstraksi Fitur dengan AZLS	V-10
V-11. Tabel Hasil Ekstraksi Fitur dengan WUMI	V-11
V-12. Tabel Hasil Percobaan Segmentasi Kata	V-12
V-13. Tabel Detail Percobaan Hasil Segmentasi Kata	V-14

DAFTAR GAMBAR

	Halaman
II-1. Gambar Tahapan Sistem Segmentasi Kata Arab Tulisan Tangan	II-6
II-2. Gambar Tahapan Pra-proses	II-7
II-3. Gambar Contoh Hasil Pemisahan Sub-kata	II-7
II-4. Gambar Contoh Deteksi <i>Middle Region</i>	II-8
II-5. Gambar Contoh Hasil Proses <i>Thinning</i>	II-9
II-6. Gambar Ilustrasi Pencarian Histogram Vertikal <i>Modified</i>	II-9
II-7. Gambar Ilustrasi Dari Histogram Vertikal <i>Modified</i> dan Informasi yang Bisa Diperoleh	II-10
II-8. Gambar Ilustrasi Proses Deteksi Lubang Pada Suatu PSP	II-11
II-9. Gambar Ilustrasi Kombinasi PSP	II-12
II-10. Gambar Ilustrasi Proses <i>Scanning</i> Untuk Setiap PSP	II-14
II-11. Gambar Contoh Dari Pembangkitan Area RC, CC, dan SA	II-14
II-12. Gambar Ilustrasi Langkah-langkah Ekstraksi <i>Line Segments</i>	II-20
II-13. Gambar Contoh dari SZ	II-22
II-14. Gambar Contoh dari AZ	II-22
III-1. Gambar Contoh Data	III-1
III-2. Gambar Kerangka Kerja Sistem	III-3
III-3. Gambar Grafik Perbandingan Akurasi Metode Ekstraksi Fitur	III-9
III-4. Gambar Penjadwalan Tahap Menentukan Ruang Lingkup Penelitian.	III-15
III-5. Gambar Penjadwalan Tahap Menentukan Dasar Teori yang Berkaitan dengan Penelitian	III-16
III-6. Gambar Penjadwalan Tahap Menentukan Kerangka Kerja, Kriteria Pengujian, Format Data Pengujian, dan Alat Penelitian	III-16
III-7. Gambar Penjadwalan Tahap Membangun Alat Penelitian Fase Insepsi	III-17
III-8. Gambar Penjadwalan Tahap Membangun Alat Penelitian Fase Elaborasi	III-17
III-9. Gambar Penjadwalan Tahap Membangun Alat Penelitian Fase Konstruksi	III-18
III-10. Gambar Penjadwalan Tahap Membangun Alat Penelitian Fase Transisi	III-18
III-11. Gambar Penjadwalan Tahap Melakukan Pengujian Penelitian	III-19
III-12. Gambar Penjadwalan Tahap Melakukan Analisa Hasil Pengujian dan Membuat Kesimpulan	III-19
IV-1. Gambar Tahapan Sistem dari Perangkat Lunak	IV-2
IV-2. Gambar Proses Pelatihan	IV-3
IV-3. Gambar Proses Segmentasi Kata	IV-5
IV-4. Gambar Contoh Pengolahan Data Latih dan Data Uji	IV-10
IV-5. Gambar Citra Simulasi WUMI	IV-11

IV-6.	Gambar Citra untuk Simulasi Penentuan Zona	IV-15
IV-7.	Gambar Hasil dari <i>Static Zoning</i>	IV-16
IV-8.	Gambar Hasil dari <i>Adaptive Zoning</i>	IV-18
IV-9.	Gambar Citra Simulasi <i>Line Segments</i>	IV-19
IV-10.	Gambar Citra Simulasi Pembangkitan Fitur	IV-21
IV-11.	Gambar Ilustrasi Proses Komparasi Hasil Segmentasi	IV-25
IV-12.	Gambar Diagram <i>Use Case</i>	IV-26
IV-13.	Gambar Diagram <i>Activity Use Case</i> Mengelola Data Pelatihan	IV-45
IV-14.	Gambar Diagram <i>Activity Use Case</i> Mengelola <i>Dataset</i> Latih	IV-46
IV-15.	Gambar Diagram <i>Activity Use Case</i> Menambah Data Karakter	IV-47
IV-16.	Gambar Diagram <i>Activity Use Case</i> Menambah Data SA	IV-48
IV-17.	Gambar Diagram <i>Activity Use Case</i> Menghapus Data	IV-48
IV-18.	Gambar Diagram <i>Activity Use Case</i> Mengelola Data Pengujian	IV-49
IV-19.	Gambar Diagram <i>Activity Use Case</i> Mengelola <i>Dataset</i> Uji	IV-49
IV-20.	Gambar Diagram <i>Activity Use Case</i> Mengelola Data Uji	IV-50
IV-21.	Gambar Diagram <i>Activity Use Case</i> Melakukan Segmentasi Kata Manual	IV-51
IV-22.	Gambar Diagram <i>Activity Use Case</i> Melakukan Pelatihan <i>Random Forests</i>	IV-52
IV-23.	Gambar Diagram <i>Activity Use Case</i> Melakukan Segmentasi Kata	IV-53
IV-24.	Gambar Diagram <i>Activity Use Case</i> Melakukan Komparasi Hasil Segmentasi	IV-54
IV-25.	Gambar Diagram <i>Activity Use Case</i> Melihat Hasil Komparasi Hasil Segmentasi	IV-54
IV-26.	Gambar Diagram <i>Activity Use Case</i> Melihat Tingkat Akurasi Hasil Segmentasi	IV-55
IV-27.	Gambar Rancangan Menu Utama	IV-56
IV-28.	Gambar Rancangan Menu Kelola Data Latih	IV-57
IV-29.	Gambar Rancangan Menu Kelola Data Uji	IV-57
IV-30.	Gambar Rancangan Sub-menu Modifikasi PSPs	IV-58
IV-31.	Gambar Rancangan Menu Pelatihan	IV-58
IV-32.	Gambar Rancangan Sub-menu Edit Parameter	IV-59
IV-33.	Gambar Rancangan Menu Segmentasi Kata	IV-59
IV-34.	Gambar Rancangan Menu Komparasi Hasil Segmentasi	IV-60
IV-35.	Gambar Rancangan Menu Hasil	IV-60
IV-36.	Gambar Diagram Kelas Analisis Mengelola Data Pelatihan	IV-62
IV-37.	Gambar Diagram Kelas Analisis Mengelola Data Pengujian	IV-62
IV-38.	Gambar Diagram Kelas Analisis Melakukan Pelatihan <i>Random Forests</i>	IV-63
IV-39.	Gambar Diagram Kelas Analisis Melakukan Segmentasi Kata	IV-63
IV-40.	Gambar Diagram Kelas Analisis Melakukan Komparasi Hasil Segmentasi	IV-64
IV-41.	Gambar Diagram Kelas Analisis Melihat Hasil Komparasi	IV-64
IV-42.	Gambar Diagram Kelas Analisis Melihat Tingkat Akurasi Hasil Segmentasi	IV-65

IV-43. Gambar Diagram Kelas	IV-67
IV-44. Gambar Diagram <i>Sequence</i> Mengelola Data Pelatihan	IV-68
IV-45. Gambar Diagram <i>Sequence</i> Mengelola <i>Dataset</i> Latih	IV-69
IV-46. Gambar Diagram <i>Sequence</i> Menambah Data Karakter	IV-70
IV-47. Gambar Diagram <i>Sequence</i> Menambah Data SA	IV-70
IV-48. Gambar Diagram <i>Sequence</i> Menghapus Data	IV-71
IV-49. Gambar Diagram <i>Sequence</i> Mengelola Data Pengujian	IV-72
IV-50. Gambar Diagram <i>Sequence</i> Mengelola <i>Dataset</i> Uji	IV-73
IV-51. Gambar Diagram <i>Sequence</i> Mengelola Data Uji	IV-74
IV-52. Gambar Diagram <i>Sequence</i> Melakukan Segmentasi Kata Manual	IV-75
IV-53. Gambar Diagram <i>Sequence</i> Melakukan Pelatihan <i>Random Forests</i> ...	IV-76
IV-54. Gambar Diagram <i>Sequence</i> Melakukan Segmentasi Kata	IV-77
IV-55. Gambar Diagram <i>Sequence</i> Melakukan Komparasi Hasil Segmentasi ...	IV-78
IV-56. Gambar Diagram <i>Sequence</i> Melihat Hasil Komparasi Hasil Segmentasi .	IV-79
IV-57. Gambar Diagram <i>Sequence</i> Melihat Tingkat Akurasi Hasil Segmentasi ..	IV-80
IV-58. Gambar Implementasi Menu Utama	IV-84
IV-59. Gambar Implementasi Menu Kelola Data Latih	IV-84
IV-60. Gambar Implementasi Menu Kelola Data Uji	IV-85
IV-61. Gambar Implementasi Sub-menu Modifikasi PSPs	IV-85
IV-62. Gambar Implementasi Menu Pelatihan	IV-86
IV-63. Gambar Implementasi Sub-menu Edit Parameter	IV-86
IV-64. Gambar Implementasi Menu Segmentasi Kata	IV-86
IV-65. Gambar Implementasi Menu Komparasi Hasil Segmentasi	IV-87
IV-66. Gambar Implementasi Menu Hasil	IV-87
V-1. Gambar Contoh Hasil Segmentasi Kata	V-14
V-2. Gambar Grafik Perbandingan Akurasi Metode Ekstraksi	V-15
V-3. Gambar Ilustrasi Skema Penentuan Garis LS	V-17
V-4. Gambar Contoh Data Gagal Segmentasi	V-18
V-5. Gambar Iterasi Proses Validasi Segmentasi (Pengujian RF)	V-19
V-6. Gambar Variasi Penulisan Karakter Arab	V-20
V-7. Gambar Redundansi Karakter Arab	V-20

DAFTAR ALGORITMA

	Halaman
II-1. Algoritma Pelatihan RF	II-17
II-2. Algoritma Pengujian RF	II-17

DAFTAR ISTILAH, SINGKATAN, DAN LAMBANG

ACD : *Average Character Density*
AZ : *Adaptive Zoning*
ACW : *Average Character Width*
AHS : *Arabic Heuristic Segmentation*
BP : *Back-propagation*
CART : *Classification and Regression Tree*
CC : *Center Character*
FFNN : *Feed Forward Neural Network*
GMI : *Geometric Moment Invariants*
MDF : *Modified Direction Features*
LS : *Line Segments-based*
OCR : *Optical Character Recognition*
PD : *Pixel Density-based*
PSP : *Prospective Segmentation Point*
RC : *Right Character*
RF : *Random Forests*
SA : *Segmentation Area*
SP : *Segmentation Point*
SZ : *Static Zoning*
UMI : *United Moment Invariants*
WBS : *Work Breakdown Structure*
WCM : *Weighted Central Moment*
WUMI : *Weighted United Moment Invariants*
ZC : *Zone Characteristics-based*

BAB I

PENDAHULUAN

1.1 Pendahuluan

Bab ini akan membahas mengenai penelitian yang akan dilakukan. Adapun yang menjadi pembahasan ialah latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan.

1.2 Latar Belakang

Pengenalan karakter atau biasa disebut dengan OCR (*Optical Character Recognition*) ialah suatu kegiatan yang bertujuan untuk mendeteksi dan mengenali karakter-karakter yang umumnya bersumber dari suatu citra (Rao, Sastri, S Chakravarthy, dan Kalyanchakravathi, 2016). Karakter yang dapat dikenali dari suatu sistem OCR dapat berupa karakter *printed* ataupun karakter *handwritten* (Bhatia, 2014). Karakter *printed* ialah karakter hasil dari mesin cetak, sedangkan karakter *handwritten* merupakan karakter hasil tulisan tangan manusia.

Ada dua pendekatan dalam melakukan pengenalan karakter, baik *printed* maupun *handwritten*, yaitu *segmentation-free* dan *segmentation-based*. Pada pendekatan *segmentation-free*, kata atau sub-kata langsung dikenali sebagai satu-kesatuan, namun pada *segmentation-based*, pengenalan dilakukan setelah kata atau sub-kata dipisahkan menjadi karakter-karakter (Maliki, Jassim, Al-Jawad, dan Sellahewa, 2012). Pada sistem OCR yang menerapkan *segmentation-based*, tentunya proses segmentasi menjadi krusial, sebab karakter yang tidak disegmentasi

dengan tepat, besar kemungkinan akan menyebabkan hasil pengenalan yang salah (Alginahi, 2013).

Sudah cukup banyak penelitian dilakukan pada ranah pengenalan karakter tulisan tangan, namun masih didapati hasil pengenalan serta segmentasi kata yang kurang memuaskan, khususnya karakter Arab. Hal ini dikarenakan sifat karakternya yang *semi-cursive*, variasi bentuk karakternya, dan gaya penulisan masing-masing orang yang berbeda (Abdeen, Afifi, & El-Sisi, 2015a). Beberapa metode yang telah digunakan untuk melakukan segmentasi kata dalam aksara Arab tulisan tangan yaitu : 1) *stroke features* (Abdulla, Al-Nassiri, dan Salam, 2008); 2) *neuro-heuristic* (Hamid dan Haraty, 2001); 3) *statistical distribution* (Alkhateeb, Jiang, Ren, dan Ipson, 2009); 4) *multi-agents* (Elnagar dan Bentrica, 2012); dan 5) *grapheme-based* (Eraqi dan Abdelazeem, 2012). Namun dari metode-metode yang telah disebutkan ada beberapa kekurangan, yaitu masalah *overlapping*, dan teknik pemisahan sub-kata yang kurang mangkus.

Abdeen, Afifi, dan El-Sisi (2015a, 2015b) dalam penelitiannya, mencoba memperbaiki masalah di atas dengan mengembangkan metode *over segmentation* dan validasi segmentasi (perbaikan hasil *over segmentation*) yang digunakan oleh Al Hamad dan Zittar (2010) dengan menambahkan mekanisme pemisahan sub-kata, memperbaiki teknik *scanning*, serta menggunakan metode ekstraksi fitur dan pemvalidasi yang berbeda. Metode ekstraksi fitur yang digunakan ialah *modified direction features* (MDF) dan *static zoning line segments-based* (SZLS), sedangkan pemvalidasi yang digunakan ialah *feed forward neural networks* (FFNN) dan *random forests* (RF). Dari hasil penelitiannya didapati bahwa dengan menggunakan

SZLS sebagai pengekstraksi fitur dan RF sebagai pemvalidasi dapat meningkatkan tingkat akurasi segmentasi.

Berkaitan dengan metode ekstraksi fitur pada sistem pengenalan karakter, Gatos, Kesidis, dan Papandreou (2011) dalam penelitiannya, membandingkan penggunaan metode ekstraksi fitur berbasis lokal, yakni *static zoning* (SZ) dengan *adaptive zoning* (AZ) yang berbasis *pixel density* (PD) dan *zone characteristics* (ZC) pada sistem pengenalan karakter latin. Hasilnya menunjukkan penggunaan AZPD (*Adaptive Zoning Pixel Density-based*) menghasilkan tingkat akurasi yang tertinggi. Sedangkan Khalaf, El Emary, dan Rifai (2006) dalam penelitiannya, menggunakan metode ekstraksi fitur berbasis global, yakni fitur *moment invariants* yang ternyata dapat digunakan pada sistem pengenalan karakter tulisan tangan Arab dikarenakan sifatnya yang tahan terhadap translasi, rotasi, dan penskalaan. Adapun salah satu pengembangan dari *moment invariants* adalah WUMI (*Weighted United Moment Invariants*), yang memiliki kemampuan lebih baik daripada *moment invariants*, terutama saat digunakan pada sistem identifikasi penulis (Samsuryadi, 2013).

Berdasarkan hal tersebut, maka pada penelitian ini akan diuji pengaruh penggunaan kombinasi metode *zoning* (SZLS, SZPD, SZZC, AZLS, AZPD, dan AZZC) dan WUMI sebagai pengekstraksi fitur terhadap tingkat akurasi sistem segmentasi kata dalam aksara Arab tulisan tangan yang dibangun berdasarkan metode yang digunakan oleh Abdeen, Afifi, dan El-Sisi (2015b), yaitu *over segmentation* sebagai pembangkit titik segmentasi dan validasi segmentasi yang bertujuan untuk memperbaiki hasil segmentasi dari proses *over segmentation*.

1.3 Rumusan Masalah

Rumusan masalah pada penelitian ini adalah pengaruh penggunaan SZLS, SZPD, SZZC, AZLS, AZPD, AZZC, dan WUMI sebagai pengekstraksi fitur pada validasi segmentasi terhadap tingkat akurasi sistem segmentasi kata tulisan tangan dalam aksara Arab.

Pertanyaan penelitian yang diperlukan untuk menjawab permasalahan, adalah sebagai berikut :

1. Bagaimana pengembangan perangkat lunak untuk melakukan segmentasi kata tulisan tangan dalam aksara Arab menggunakan SZLS, SZPD, SZZC, AZLS, AZPD, AZZC, dan WUMI sebagai pengekstraksi fitur untuk validasi terhadap hasil segmentasi.
2. Bagaimana mencari tingkat akurasi dari SZLS, SZPD, SZZC, AZLS, AZPD, AZZC, dan WUMI sebagai pengekstraksi fitur untuk validasi terhadap hasil segmentasi pada sistem segmentasi kata tulisan tangan dalam aksara Arab.

1.4 Tujuan Penelitian

Tujuan penelitian ini adalah :

1. Mengembangkan perangkat lunak untuk melakukan segmentasi kata tulisan tangan dalam aksara Arab menggunakan SZLS, SZPD, SZZC, AZLS, AZPD, AZZC, dan WUMI sebagai pengekstraksi fitur untuk validasi terhadap hasil segmentasi.

2. Mencari tingkat akurasi dari SZLS, SZPD, SZZC, AZLS, AZPD, AZZC, dan WUMI sebagai pengekstraksi fitur untuk validasi terhadap hasil segmentasi pada sistem segmentasi kata tulisan tangan dalam aksara Arab.

1.5 Manfaat Penelitian

Manfaat penelitian ini ialah :

1. Perangkat lunak yang dihasilkan dapat digunakan untuk melakukan segmentasi kata tulisan tangan dalam aksara Arab menggunakan metode *over segmentation* dan validasi segmentasi dengan pengekstraksi fitur SZLS, SZPD, SZZC, AZLS, AZPD, AZZC, dan WUMI.
2. Menjadi referensi berkaitan dengan pemilihan metode ekstraksi fitur dalam ranah pengenalan karakter Arab tulisan tangan.

1.6 Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut :

1. Data yang digunakan (baik data latih maupun data uji) berupa citra diam.
2. Data latih dan data uji yaitu karakter dan kata Arab hasil tulisan tangan.
3. Sistem tidak mengatasi *ligatures* (karakter yang bergabung secara vertikal).

1.7 Sistematika Penulisan

Sistematika penulisan dari skripsi ini adalah sebagai berikut :

1. Bab I Pendahuluan

Bab ini berisi pembahasan mengenai latar belakang penelitian, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan.

2. Bab II Kajian Literatur

Bab ini berisi teori-teori yang digunakan untuk memahami permasalahan yang dibahas pada penelitian ini.

3. Bab III Metodologi Penelitian

Bab ini berisi deskripsi data penelitian, tahapan-tahapan yang akan dilaksanakan pada penelitian, metode pengembangan perangkat lunak, serta manajemen dari penelitian ini.

4. Bab IV Pengembangan Perangkat Lunak

Bab ini berisi pembahasan proses pengembangan perangkat lunak segmentasi kata Arab tulisan tangan yang terdiri dari proses analisis kebutuhan, perancangan, implementasi, serta pengujian perangkat lunak.

5. Bab V Hasil dan Analisa Penelitian

Bab ini berisi penjelasan mengenai hasil percobaan penelitian yang telah dilakukan beserta analisa terhadap hasil percobaan tersebut.

6. Bab VI Kesimpulan dan Saran

Bab ini berisi kesimpulan dari hasil penelitian yang diperoleh beserta saran terhadap penelitian ini untuk kedepannya.

1.8 Kesimpulan

Pada bab ini telah dibahas mengenai penelitian yang akan dilakukan, yaitu mengetahui pengaruh dari penggunaan SZLS, SZPD, SZZC, AZLS, AZPD, AZZC, dan WUMI sebagai pengekstraksi fitur pada validasi segmentasi terhadap tingkat akurasi sistem segmentasi kata tulisan tangan dalam aksara Arab. Selanjutnya, teori-teori yang berkaitan dengan penelitian akan dibahas pada bab II.

DAFTAR PUSTAKA

- Abdeen, R. M., Afifi, A., & El-Sisi, A. B. (2015a). An Enhanced Technique for Offline Arabic Handwritten Words Segmentation. In A. Gelbukh(Ed): *CICLing* (Vol. Part II, pp. 663–681).
- Abdeen, R. M., Afifi, A., & El-Sisi, A. B. (2015b). Improved Arabic Handwriting Word Segmentation Approach using Random Forests. *2015 IEEE/ACS 12th International Conference of Computer Systems and Applications (AICCSA)*. <https://doi.org/10.1109/AICCSA.2015.7507105>
- Abdulla, S., Al-Nassiri, A., & Salam, R. . (2008). Offline Arabic Handwriting Word Segmentation Using Rotational Invariant Segments Features. *The International Arab Journal of Information Technology*, 5(2).
- Al Hamad, H. A., & Zittar, A. R. (2010). Development of an efficient neural-based segmentation technique for Arabic handwriting recognition. *Pattern Recognition*, 43(8), 2773–2798. <https://doi.org/10.1016/j.patcog.2010.03.005>
- Alginahi, Y. M. (2013). A survey on Arabic character segmentation. *International Journal of Document Analysis and Recognition*, 16(12). <https://doi.org/10.1007/s10032-012-0188-6>
- Alkhateeb, J. ., Jiang, K., Ren, J., & Ipson, S. . (2009). Component-based Segmentation of Words from Handwritten Arabic Text. *International Journal of Computer Systems Science and Engineering*, 5(1).
- Althobaiti, H., & Lu, C. (2017). A Survey on Arabic Optical Character Recognition and an Isolated Handwritten Arabic Character Recognition Algorithm using Encoded Freeman Chain Code, 2–7.
- Bhatia, E. N. (2014). Optical Character Recognition Techniques : A Review. *International Journal of Advanced Research in Computer Science and Software Engineering*, 4(5), 1219–1223.
- Breiman, L., & Cutler, A. (n.d.). Random forests - classification description. Retrieved November 21, 2017, from https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm
- Breiman, L., & Cutler, A. (2014). *Random Forests for Beginners*. Salford Systems 2014.
- Dahi, M., Semaary, N. A., & Hadhoud, M. M. (2015a). *A Comparative Study of Different Approaches of Primitive Printed Arabic Optical Character Recognition*. Shebin EI-Kom, Egypt.
- Dahi, M., Semaary, N. A., & Hadhoud, M. M. (2015b). Primitive Printed Arabic Optical Character Recognition using Statistical Features. In *IEEE Seventh International Conference on Intelligent Computing and Information Systems* (pp. 567–571).
- Dharma, K. S. (2016). *Pengenalan Tulisan Tangan Angka Cina Menggunakan Weighted United Moment Invariant dan Self Organizing Maps*. Universitas Sriwijaya.
- Elnagar, A., & Bentrica, R. (2012). A Multi-Agent Approach to Arabic Handwritten Text Segmentation. *Journal of Intelligent Learning System and Applications*.

- Eraqi, H. ., & Abdelazeem, S. (2012). A new Efficient Graphemes Segmentation Technique for Offline Arabic Handwriting. In *International Conference on Frontiers in Handwriting Recognition*.
- Gatos, B., Kesidis, A. L., & Papandreou, A. (2011). Adaptive Zoning Features for Character and Word Recognition. In *International Conference on Document Analysis and Recognition*. <https://doi.org/10.1109/ICDAR.2011.234>
- Hamid, A., & Haraty, R. (2001). A Neuro-Heuristic Approach for Segmenting Handwritten Arabic Text. *IEEE*.
- Impedovo, D., & Pirlo, G. (2013). Zoning methods for handwritten character recognition : A survey. *Pattern Recognition*, 1–13. <https://doi.org/10.1016/j.patcog.2013.05.021>
- Isheawy, N. A. M., & Hasan, H. (2015). Optical Character Recognition (OCR) System. *IOSR Journal of Computer Engineering (IOSR-JCE)*, 17(2), 22–26. <https://doi.org/10.9790/0661-17222226>
- Khalaf, K., El Emary, I. M. M., & Rifai, B. (2006). Probabilistic Artificial Neural Network For Recognizing the Arabic Hand Written Characters. *Journal of Computer Science*, 2(12). <https://doi.org/10.3844/jcssp.2006.879.884>
- Kruchten, P. (2000). *The Rational Unified Process An Introduction* (2nd ed.). Addison Wesley.
- Maliki, M., Jassim, S., Al-Jawad, N., & Sellahewa, H. (2012). Arabic Handwritten : Pre-Processing and segmentation. In *SPIE*. <https://doi.org/10.1117/12.917555>
- Nasution, A. (2017). Abjad, Aksara, Alfabet, Huruf, dan Karakter. Bedanya Apa? Retrieved November 17, 2017, from <https://albadr.blog/2013/05/31/abjad-aksara-alfabet-huruf-dan-karakter-bedanya-apa/>
- Pamungkas, R. P., & Shamsuddin, S. M. (2009). Weighted Central Moment for Pattern Recognition: Derivation, Analysis of Invariancess, and Simulation using Letter Characters. *Third Asia International Conference on Modelling & Simulation on IEEE*, 102–106.
- Polamuri, S. (2017). How the random forest algorithm works in machine learning. Retrieved January 1, 2017, from <http://dataaspirant.com/2017/05/22/random-forest-algorithm-machine-learning/>
- Rao, N. V., Sastri, A. S., S Chakravarthy, A., & Kalyanchakravathi, A. (2016). OPTICAL CHARACTER RECOGNITION TECHNIQUE ALGORITHMS. *Journal of Theoretical and Applied Information Technology*, 83(2).
- Samsuryadi. (2013). *Biomimetic Pattern Recognition for Writer Identification Using Geometrical Moment Functions*. Universiti Teknologi Malaysia.