

**SEGMENTASI MANUSKRIP BERAKSARA JAWA
MENGUNAKAN PROFIL PROYEKSI DAN KOMPONEN
TERKONEKSI (CONNECTED COMPONENT)**

Skripsi



Diajukan oleh :

Samuel Eddijanto Harjono

71150044

PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA
YOGYAKARTA

2019

**SEGMENTASI MANUSKRIP BERAKSARA JAWA
MENGUNAKAN PROFIL PROYEKSI DAN KOMPONEN
TERKONEKSI (CONNECTED COMPONENT)**

Skripsi



Diajukan kepada Fakultas Teknologi Informasi Program Studi Informatika
Universitas Kristen Duta Wacana
Sebagai salah satu syarat dalam memperoleh gelar Sarjana Komputer

Disusun oleh :

SAMUEL EDDIJANTO HARJONO

71150044

PROGRAM STUDI INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA YOGYAKARTA

2019

PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul:

SEGMENTASI MANUSKRIP BERAKSARA JAWA MENGGUNAKAN PROYEKSI PROFIL DAN KOMPONEN TERKONEKSI (CONNECTED COMPONENT)

yang saya kerjakan untuk melengkapi sebagian persyaratan menjadi Sarjana Komputer pada pendidikan Sarjana Program Studi Informatika Fakultas Teknologi Informasi Universitas Kristen Duta Wacana, bukan merupakan tiruan atau duplikasi dari skripsi keserjanaan di lingkungan Universitas Kristen Duta Wacana maupun di Perguruan Tinggi atau instansi manapun, kecuali bagian yang sumber informasinya dicantumkan sebagaimana mestinya.

Jika dikemudian hari didapati bahwa hasil skripsi ini adalah hasil plagiasi atau tiruan dari skripsi lain, saya bersedia dikenai sanksi yakni pencabutan gelar keserjanaan saya.

Yogyakarta, 7 Januari 2020



SAMUEL EDDIJANTO HARJONO
71150044

HALAMAN PERSETUJUAN

Judul Skripsi : SEGMENTASI MANUSKRIP BERAKSARA JAWA
MENGUNAKAN PROYEKSI PROFIL DAN
KOMPONEN TERKONEKSI (CONNECTED
COMPONENT)

Nama Mahasiswa : SAMUEL EDDIJANTO HARJONO

N I M : 71150044

Matakuliah : Skripsi (Tugas Akhir)

Kode : TIW276

Semester : Gasal

Tahun Akademik : 2019/2020

© UKDW
Telah diperiksa dan disetujui di
Yogyakarta,
Pada tanggal 7 Januari 2020

Dosen Pembimbing I



Aditya Wikan Mahastama, S.Kom.,
M.Cs.

Dosen Pembimbing II



Lucia Dwi Krisnawati, Dr. Phil.

HALAMAN PENGESAHAN

SEGMENTASI MANUSKRIP BERAKSARA JAWA MENGGUNAKAN PROYEKSI PROFIL DAN KOMPONEN TERKONEKSI (CONNECTED COMPONENT)

Oleh: SAMUEL EDDIJANTO HARJONO / 71150044

Dipertahankan di depan Dewan Penguji Skripsi
Program Studi Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana - Yogyakarta
Dan dinyatakan diterima untuk memenuhi salah satu syarat memperoleh gelar
Sarjana Komputer
pada tanggal 11 Desember 2019

Yogyakarta, 7 Januari 2020
Mengesahkan,

Dewan Penguji:


1. Aditya Wikan Mahastama, S.Kom., M.Cs.
2. Lucia Dwi Krisnawati, Dr. Phil.
3. Lukas Chrisantyo, S.Kom., M.Eng.
4. Maria Nila Anggia Rini, S.T, M.T.I

DUTA WACANA

Dekan

Ketua Program Studi


(Restyandito, S.Kom., MSIS., Ph.D.)


(Gloria Virginia, Ph.D.)

UCAPAN TERIMA KASIH

Penulis mengucapkan Puji dan Syukur kepada Tuhan yang Mahaesa atas karuniaNya sehingga laporan tugas akhir berjudul “Segmentasi Manuskrip Beraksara Jawa Menggunakan Profil proyeksi dan Komponen Terkoneksi (Connected Component)” dapat selesai tanpa kurang suatu apapun. Laporan ini bertujuan untuk menyelesaikan studi tingkat Strata – 1 di program studi Informatika Universitas Kristen Duta Wacana. Penulis berharap laporan ini dapat berguna baik bagi pembaca maupun penulis.

Dalam menyelesaikan penelitian dan penulisan laporan ini, penulis mendapat banyak bantuan baik secara langsung maupun tidak langsung. Oleh karena itu, pada kesempatan ini penulis menyampaikan terimakasih kepada:

1. Bapak Aditya Wikan Mahastama S.Kom., M.Cs. selaku dosen pembimbing I yang telah membimbing dan mengarahkan penulis dalam menyelesaikan tugas akhir.
2. Ibu Dr. Phil. Lucia Dwi Krisnawati selaku dosen pembimbing II yang telah membimbing dan mengarahkan penulis dalam menyelesaikan tugas akhir.
3. Keluarga penulis yang telah mendukung baik dalam doa maupun dukungan secara moral dan material.
4. Teman-teman penulis yang tidak dapat disebutkan satu per satu dalam tulisan ini.

Penulis menyadari laporan ini masih jauh dari kata sempurna. Oleh karena itu, segala saran dan kritik yang disampaikan akan dipertimbangkan dan diharapkan dapat menjadi koreksi baik bagi penulis maupun pihak lain yang terlibat. Akhir kata, penulis sampaikan terimakasih dan semoga laporan ini dapat digunakan sebagaimana mestinya.

Yogyakarta, November 2019

Samuel Eddijanto H.

Penulis

DAFTAR ISI

PERNYATAAN KEASLIAN SKRIPSI.....	iii
HALAMAN PERSETUJUAN.....	iv
UCAPAN TERIMA KASIH.....	vi
INTISARI.....	vii
DAFTAR ISI.....	viii
DAFTAR GAMBAR	x
DAFTAR TABEL.....	x
BAB I.....	1
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah.....	2
1.4 Tujuan Penelitian.....	2
1.5 Manfaat Penelitian.....	2
1.6 Metodologi Penelitian	3
1.7 Sistematika Penulisan.....	4
BAB II Tinjauan Pustaka dan Landasan Teori	5
2.1 Tinjauan Pustaka	5
2.1.1 Segmentasi Baris untuk Dokumen Sunda Lampau	5
2.1.2 Segmentasi Aksara Bali dengan Profil proyeksi	6
2.1.3 Ekstraksi dan Pelabelan Komponen Terkoneksi dari Citra Kromosom.....	6
2.1.4 Teknik Preprocessing dalam Pengenalan sebuah Karakter	7
2.1.5 Usulan Model untuk Transliterasi Aksara Jawa.....	8
2.2 Landasan Teori	9
2.2.1 Aksara Jawa.....	9
2.2.2 Segmentasi Citra	10
2.2.3 Komponen Terkoneksi (Connected Component).....	11
2.2.4 Profil proyeksi	12
2.2.5 Pelabelan	13

BAB III PERANCANGAN SISTEM	14
3.1 Perancangan Sistem.....	14
3.2 Blok Diagram Sistem	14
3.3 Analisis Kebutuhan Sistem	15
3.3.1 Kebutuhan Fungsional.....	16
3.3.2 Kebutuhan Non-Fungsional	16
3.4 Perancangan Antarmuka	17
3.5 Rencana Pengujian Skenario Sistem	18
BAB IV IMPLEMENTASI dan Analisis Sistem.....	19
4.1 Implementasi Sistem	19
4.1.1 Thresholding.....	19
4.1.2 Perhitungan Histogram Proyeksi Vertikal.....	20
4.1.3 Pendeteksian Karakter Utama dan Pemotongan baris.....	20
4.1.4 Komponen Terkoneksi	22
4.1.5 Pemotongan Aksara.....	22
4.2 Hasil Pengujian dan Analisis Sistem.....	23
4.3 Kekurangan Sistem	44
BAB V Kesimpulan dan Saran	45
5.1 Kesimpulan.....	45
5.2 Saran	46
DAFTAR PUSTAKA	47
LAMPIRAN A	48
LAMPIRAN B	54

DAFTAR GAMBAR

<i>Gambar 4.1</i> Hasil Pemisahan Foreground dan Background menggunakan Threshold	19
<i>Gambar 4.2</i> Pseudocode Proyeksi Vertikal (Widiarti, 2007)	20
<i>Gambar 4.3</i> Pseudocode Deteksi Baris (Widiarti, 2007).....	21
<i>Gambar 4.4</i> Hasil dari Pendeteksian Baris	21
<i>Gambar 4.5</i> Hasil Segmentasi Baris yang Pertama	21
<i>Gambar 4.6</i> Hasil Segmentasi Baris yang Kedua.....	22
<i>Gambar 4.7</i> Hasil Segmentasi Baris yang Ketiga.....	22
<i>Gambar 4.8</i> Pseudocode Pemotongan Aksara (Widiarti, 2007).....	23
<i>Gambar 4.9</i> Contoh Citra Inputan	24
<i>Gambar 4.10</i> Citra Hasil Profil proyeksi.....	24
<i>Gambar 4.11</i> Citra Hasil Komponen Terkoneksi yang Berhasil Dikenali	25
<i>Gambar 4.12</i> Citra Hasil Komponen Terkoneksi yang Gagal Dikenali (Over Segmented)	25
<i>Gambar 4.13</i> Hasil Segmentasi Profil Proyeksi dengan Nilai Ambang Batas 140	27
<i>Gambar 4.14</i> Hasil Segmentasi Profil Proyeksi dengan Nilai Ambang Batas 210	28
<i>Gambar 4. 15</i> Cintra Inputan 08.JPG	30
<i>Gambar 4. 16</i> Citra Inputan 09.JPG	31
<i>Gambar 4. 17</i> Citra Inputan 12.JPG	31
<i>Gambar 4. 18</i> Citra Inputan 01.JPG	32
<i>Gambar 4. 19</i> Citra Inputan 14.JPG	32
<i>Gambar 4. 20</i> Profil Proyeksi dengan Ambang Batas 140.....	33
<i>Gambar 4. 21</i> Profil Proyeksi dengan Ambang Batas 210.....	34
<i>Gambar 4. 22</i> Komponen Terkoneksi dengan Ambang Batas 140 pada Baris Pertama gambar 08.JPG.....	35
<i>Gambar 4. 23</i> Komponen Terkoneksi dengna Ambang Batas 210 pada Baris Pertama gambar 08.JPG.....	35
<i>Gambar 4. 24</i> Profil Proyeksi dengan Ambang Batas 140.....	35
<i>Gambar 4. 25</i> Profil Proyeksi dengan Ambang Batas 210.....	36
<i>Gambar 4. 26</i> Komponen Terkoneksi dengan Ambang Batas 140 pada baris 11 gambar 09.JPG.....	36
<i>Gambar 4. 27</i> Komponen Terkoneksi dengan Ambang Batas 210 pada Baris 11 gambar 09.JPG.....	37
<i>Gambar 4. 28</i> Profil Proyeksi dengan Ambang Batas 140.....	37
<i>Gambar 4. 29</i> Profil Proyeksi dengan Ambang Batas 210.....	38

Gambar 4. 30 Komponen Terkoneksi dengan Ambang Batas 140 pada baris empat belas gambar 01.JPG	38
Gambar 4. 31 Komponen Terkoneksi dengan Ambang Batas 210 pada baris empat belas gambar 01.JPG	39
Gambar 4. 32 Profil Proyeksi dengan Ambang Batas 140.....	39
Gambar 4. 33 Profil Proyeksi dengan Ambang Batas 210.....	40
Gambar 4. 34 Komponen Terkoneksi dengan Ambang Batas 140 pada Baris 20 gambar 12.JPG.....	41
Gambar 4. 35 Komponen Terkoneksi dengan Ambang Batas 210 pada Baris 20 gambar 12.JPG.....	41
Gambar 4. 36 Profil Proyeksi dengan Ambang Batas 140.....	42
Gambar 4. 37 Profil Proyeksi dengan Ambang Batas 210.....	42
Gambar 4. 38 Komponen Terkoneksi dengan Ambang Batas 140 pada Baris Empat gambar 14.JPG.....	43
Gambar 4. 39 Komponen Terkoneksi dengan Ambang Batas 210 pada Baris empat gambar 14.JPG.....	43

© UKDW

DAFTAR TABEL

Tabel 4. 1 Perbandingan Pemotongan Citra dengan Nilai Ambang Batas 170	25
Tabel 4. 2 Hasil Pemotongan Citra dengan Beberapa Nilai Ambang Batas	26
Tabel 4. 3 Hasil Pemotongan Aksara Jawa Setelah Diamati	28
Tabel 4. 4 Tabel Akurasi Setiap Nilai Ambang Batas	29

©UKDW

BAB I

PENDAHULUAN

1.1 Latar Belakang

Manuskrip Jawa adalah sebuah bukti bersejarah yang ada di Indonesia, karena dalam manuskrip Jawa kebanyakan dikisahkan mengenai kehidupan dari zaman kerajaan yang pernah berdiri di Indonesia. Tetapi sebuah manuskrip tidak bisa bertahan lama karena manuskrip tersebut dituliskan pada sebuah kertas maupun kulit pohon, sehingga semakin lama umur manuskrip tersebut semakin renta juga media penulisannya dan pasti akan mengalami kerusakan.

Dengan berkembangnya teknologi saat ini, digitalisasi adalah cara terbaik untuk mengubah sebuah sumber sejarah berbentuk manuskrip menjadi sebuah file yang bisa dibuka kapanpun dan di manapun. Dalam proses digitalisasi tersebut diperlukan pengenalan karakter yang tertulis pada manuskrip tersebut. Selain menyeleksi naskah asli dari derau, proses segmentasi adalah bagian kunci dari penelitian ini karena dalam proses ini dilakukan pembacaan posisi dan letak suatu karakter yang sudah dipisahkan antara *background* dan *foreground* serta dilakukan pemotongan untuk setiap karakter yang telah diketahui lokasinya. Keberhasilan dari proses segmentasi sendiri bergantung pada cara aksara Jawa dituliskan dalam manuskrip yang diteliti, misalnya ukuran karakter, dan gaya penulisannya.

Hasil dari proses Segmentasi adalah aksara Jawa yang sudah terpisah antara aksara yang satu dengan yang lain, dan dalam hasil segmentasi tersebut, aksara Jawa sandhangan dan pasangan sudah menjadi satu bagian bersama dengan aksara nglegena. Akan tetapi untuk mendapatkan hasil tersebut tidaklah mudah, akan ada hambatan yang bisa mempengaruhi proses segmentasi tersebut. Contohnya seperti miringnya manuskrip Jawa pada buku yang akan digunakan sebagai dokumen inputan, adanya lubang pada manuskrip Jawa, dan hilangnya tulisan aksara Jawa pada dokumen tersebut.

1.2 Rumusan Masalah

1. Bagaimana metode profil proyeksi dan komponen terkoneksi dapat melakukan proses segmentasi citra teks beraksara Jawa secara optimal?

1.3 Batasan Masalah

Batasan masalah dari permasalahan yang akan diteliti adalah :

1. Input berupa citra dokumen dengan format .jpg atau .jpeg yang diambil dari data latih milik TRAWACA (Proyek kerjasama WIKIMEDIA dan UKDW)
2. Citra Dokumen harus bersih dari noktah dan tidak miring
3. Dokumen berupa manuskrip tercetak beraksara Jawa yang memiliki ketebalan berbeda

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah Membuat sebuah sistem yang membantu proses digitalisasi manuskrip dengan cara melakukan proses segmentasi atau memotong aksara Jawa kedalam bagian-bagian kecil yang mudah dikenali agar dapat dilakukan ekstraksi fitur untuk proses pelabelan di tahap penelitian selanjutnya.

1.5 Manfaat Penelitian

Manfaat yang dapat diperoleh manuskrip beraksara Jawa dapat dilestarikan dan dapat membantu museum dalam mengenalkan suatu manuskrip beraksara Jawa kepada pengunjung museum melalui gambar manuskrip aksara Jawa yang sudah tersegmentasi.

1.6 Metodologi Penelitian

Penelitian ini dilakukan melalui beberapa tahapan yakni :

1. Studi Literatur

Dilakukan dengan cara mempelajari teori–teori yang diperlukan dalam penelitian melalui artikel–artikel yang berkaitan dengan metode *Connected Component* dan Profil proyeksi.

2. Tahap analisis kebutuhan sistem.

Sistem yang akan dibuat merupakan sebuah system yang akan membantu dalam hal segmentasi citra manuskrip beraksara Jawa yang akan dirancang menggunakan bahasa pemrograman Python 3.7.0

3. Tahap pembuatan sistem.

Sistem akan dibuat dengan sistem aplikasi yang berfungsi untuk mengolah inputan citra manuskrip kemudian dilakukan pemisahan antara *foreground* dengan *background* citra atau yang dikenal dengan binerisasi menggunakan *thresholding* , lalu citra akan dilakukan proses segmentasi menggunakan metode profil proyeksi vertikal dan horizontal. Namun, jika hasilnya terdapat beberapa aksara Jawa yang belum terpisah dengan baik akan dilanjutkan segmentasi iterasi kedua yaitu menggunakan segmentasi *connected component* , dari proses tersebut akan didapatkan beberapa label berisikan aksara Jawa yang sudah dibedakan antara satu dengan yang lain, kemudian label tersebut dijadikan sebagai acuan untuk melakukan proses pemotongan setiap aksara, sandhangan, dan pasangan berdasarkan baris dan kolom yang sudah terlabeli.

4. Tahap pengujian sistem.

Pengujian dilakukan dengan mencermati dan menghitung suatu tingkat keberhasilan sistem yang dibuat pada citra keluaran yang sudah selesai dipotong. Jika citra keluaran memiliki persentase keberhasilan tinggi dalam pemotongan karakter, maka sistem yang dibuat terbilang berhasil.

1.7 Sistematika Penulisan

BAB I PENDAHULUAN, berisi tentang latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, metodologi penelitian dan sistematika penulisan.

BAB II LANDASAN TEORI, berisi tentang tinjauan pustaka dan landasan teori yang digunakan sebagai teori pendukung di dalam penelitian ini.

BAB III PERANCANGAN SISTEM, berisi tentang penjelasan dan rancangan sistem yang akan dibuat.

BAB IV IMPLEMENTASI DAN UJI ANALISIS SISTEM, membahas tentang penerapan sistem dan menguji sistem tersebut untuk dianalisis.

BAB V KESIMPULAN, berisi tentang kesimpulan beserta saran yang akan diberikan peneliti.

© UKDW

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil analisa dan pembahasan pada Bab 3 dan Bab 4 serta hasil dari simulasi sistem dapat disimpulkan sebagai berikut:

1. Sistem sudah cukup berhasil untuk melakukan proses pemotongan citra beraksara Jawa menjadi beberapa karakter sesuai dengan tujuan penelitian, baik dalam pemotongan baris maupun pemotongan aksara Jawa menggunakan metode komponen terkoneksi dan profil proyeksi
2. Sebuah kualitas ketebalan citra masukan dan nilai ambang batas (*Threshold*) prapemrosesan ternyata memiliki pengaruh yang besar terhadap proses pemotongan karakter beraksara Jawa. Semakin tinggi kualitas ketebalan citra masukan dan semakin tinggi nilai ambang batas prapemrosesan, dan ambang batas yang digunakan bernilai 140, 170, 200, 210 dengan hasil terbaik pada ambang batas 210.
3. Dari kedua metode ternyata persentase keakuratan dalam melakukan segmentasi aksara Jawa antara metode profil proyeksi dan komponen terkoneksi didapatkan bahwa profil proyeksi memiliki keakuratan yang lebih tinggi dari komponen terkoneksi dengan nilai untuk profil proyeksi terendah 27.55% dan tertinggi 69.84% sedangkan untuk komponen terkoneksi terendah 17.97% dan tertinggi 50%.
4. Kelemahan dari sistem ini adalah pemotongan karakter aksara Jawa masih menghasilkan potongan citra yang *Over Segmented* sejumlah 56 potong untuk metode komponen terkoneksi dan 19 potong untuk profil proyeksi walaupun nilai ambang batas yang digunakan menghasilkan pemotongan yang bagus , terutama jika sebuah citra masukan merupakan citra teks beraksara Jawa yang saling berhimpitan baik dari segi baris maupun kolom untuk setiap aksaranya.

5.2 Saran

Sistem ini sangat memungkinkan untuk dilakukan pengembangan lebih lanjut. Saran yang diajukan penulis dalam pengembangan sistem kedepannya adalah sebagai berikut:

1. Dalam proses pemilihan citra masukan diharuskan untuk memilih citra yang benar-benar bagus terutama cetakan aksara Jawa pada citra tersebut.
2. Pada penelitian selanjutnya dapat ditambahkan sebuah metode binerisasi menggunakan *Adaptive Thresholding* agar dihasilkan citra keluaran yang optimal.
3. Dan dapat dilakukan analisis mendalam untuk proses penyusunan ulang potongan karakter teks beraksara Jawa pada metode komponen terkoneksi.

©UKDW

DAFTAR PUSTAKA

- Asano, T., Tanaka, H. (2010). *In-place Algorithm for Connected Components Labelling*. Journal of Pattern Recognition Research 1 (JPARR).
- Alginahi, Yasser (2010). *Preprocessing Techniques in Character Recognition* .
DOI: 10.5772/9776 · Source: InTech
- Chokkalingam, A., Rajaraman, S. (2013). *Connected Components Labelling and Extraction Based Interphase Removal from Chromosome Images*. *International Journal of Bio-Science and Bio-Technology*. Vol.5 , No.1
- Gonzalez, R., Woods. R. (2008). *Digital Image Processing (Third edition)*. United State of America. Pearson Education.
- Hendry, J. (2011). *Using Profile Projection to Segment Character In Image*.
- Hadi, S. , Hidayat, A., Paulus, M. , Suryani, M. (2017). *Investigasi Segmentasi Baris Untuk Citra Dokumen Sunda Lampau*. *Jurnal Informatika dan Komputer (JIKO)*. Vol.2, No.2
- Ingold, R., Zramdini, A. (1993). *Optical Font Recognition from Projection Profiles*. Electronic Publishing, Vol. 6 : 249-260
- Kesiman, P., Sunarya, G. (2015). *Segmentasi Citra Tulisan Tangan Aksara Bali Berbasis Proyeksi Vertikal Dan Horizontal*. *Jurnal Informatika* Vol.9, No. 1
- Mohamed, N. (2001). *Aksara Jawi: Makna dan Fungsi*.
- Gonzalez, R., Woods. R. (2008). *Digital Image Processing (Third edition)*. 121-131.
- Widiarti, A. (2013). *Line Segmentation Of Javanese Image Of Manuscripts in Javanese Scripts*. *International Journal of Engineering Innovation & Research*. Vol.2, No.3.
- Widiarti, A., Pulungan, R., Harjoko, A., Marsono, Hartati S.(2018). *A Proposed Model for Javanese Manuscript Images Transliteration*. *IOP Conf. Series: Journal of Pyhsics: Conf. Series* 1098 (2018) 012014. doi :10.1088/1742-6596/1098/1/012014