

**PENGENALAN KARAKTER KOREA (HANGEUL) TULISAN
TANGAN MENGGUNAKAN *ZONE CENTROID AND ZONE*
(ZCZ) DAN *BACKPROPAGATION***

*Diajukan Sebagai Syarat
untuk Menyelesaikan Pendidikan Program Strata-1
pada Jurusan Teknik Informatika Fakultas Ilmu Komputer UNSRI*



Oleh :

DWI MAHARANI

NIM : 09111002063

**Jurusan Teknik Informatika
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2016**

LEMBAR PENGESAHAN TUGAS AKHIR

Pengenalan Karakter Korea (Hangeul) Tulisan Tangan Menggunakan *Zone Centroid and Zone (ZCZ)* dan *Backpropagation*

Oleh :

DWI MAHARANI

NIM : 09111002063

Pembimbing I,

Palembang, Januari 2016
Pembimbing II,

Drs. Saparudin, M.T., Ph.D
NIP. 196904121995021001

Hadipurnawan Satria, Ph.D
NIPUS. 198004182015109101

Menyetujui,
Ketua Jurusan Teknik Informatika

Drs. Megah Mulya, M.T
NIP 196602202006041001

**PENGENALAN KARAKTER KOREA (HANGUL) TULISAN TANGAN
MENGUNAKAN ZONE CENTROID AND ZONE (ZCZ) DAN
BACKPROPAGATION**

Oleh:

Dwi Maharani

09111002063

ABSTRAK

Pengenalan karakter tulisan tangan dalam praktiknya telah banyak membantu dalam aktifitas kehidupan. Namun karakter dengan huruf latin lebih banyak diteliti dibanding karakter non-latin. Tidak seperti karakter latin, karakter Korea (Hangul) memiliki struktur penulisan dua arah sehingga membuat Hangul memiliki tingkat kompleksitas yang lebih tinggi dibandingkan dengan alfabet Latin yang ditulis hanya dengan satu arah. Pada penelitian ini pengenalan karakter Hangul tulisan tangan dilakukan dengan menggunakan metode *Zone Centroid and Zone (ZCZ)* sebagai metode ekstraksi ciri dan metode *Backpropagation* sebagai metode pengenalannya. Pelatihan dimulai dengan citra latih dipilih dan dilakukan serangkaian proses yaitu prapengolahan yang terdiri dari penskalaan, *grayscale*, binerisasi, dan skeletonisasi, ekstraksi ciri menggunakan *ZCZ*, dan pelatihan serta pengujian menggunakan *Backpropagation*. Pengujian menggunakan 200 data tulisan tangan dari 10 karakter yang masing-masing terdiri dari dua buah variasi penulisan yaitu tegak lurus dan miring. Proses pelatihan berhenti dengan acuan nilai MSE lebih kecil dari *error tolerance*. Hasil pengujian dengan perangkat lunak yang dikembangkan mampu mengenali tulisan tangan karakter Hangul dengan variasi berupa data tegak lurus dan data miring dengan presentase keakuratan rata-rata sebesar 100% untuk data latih dan 89% untuk data uji.

Kata kunci: Pengenalan karakter tulisan tangan, *backpropagation*, *Zone Centroid and Zone (ZCZ)*, dan Hangul.

**HANDWRITTEN KOREAN CHARACTER (HANGEUL) RECOGNITION
USING ZONE CENTROID AND ZONE (ZCZ) AND
BACKPROPAGATION**

Dwi Maharani

09111002063

ABSTRACT

Research in Handwritten character recognition is popular for various practical application in daily life activity but research in latin character is more than in non-latin character. Korean character (Hangul) is composed in two dimensional space which is horizontal and vertical while Latin is composed in one dimensional space only. This research using Zone Centroid and Zone (ZCZ) as feature extractor and backpropagation as training and testing method. Training stage is started by choosing the training data, then the data is computed in preprocessing stage such as scaling, grayscaling, thresholding, and thinning. Next, the data's feature will be obtained using Zone Centroid and Zone, so the feature could be the input of the training and testing phase using backpropagation. We obtained 100% recognition rate of training data and 89% recognition rate of testing data in italic and straight variation of data.

Keyword: Handwritten character recognition, backpropagation, Zone Centroid and Zone (ZCZ), Hangeul.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pengenalan karakter tulisan tangan memiliki peranan penting dalam pengenalan pola dan pengolahan citra. Pengenalan karakter tulisan tangan atau *Handwritten Character Recognition* (HCR) dalam praktiknya telah banyak membantu dalam aktifitas kehidupan, seperti pengembangan alat bantu baca bagi tuna netra, pengenalan tulisan pada cek bank, dan aplikasi membaca kode pin pada surat pos otomatis (Rajashekararadhya dan Ranjan, 2008). Penelitian atas pengenalan karakter tulisan tangan telah banyak dilakukan, namun karakter dengan huruf latin lebih banyak diteliti dibanding karakter non-latin (Sinha, Rani, dan Dhiret, 2012).

Tidak seperti karakter latin, karakter non-latin, seperti karakter Korea (Hangul), dibentuk oleh lebih dari satu unit karakter (*grapheme*). Dalam penulisan karakter Hangul, posisi peletakan setiap *grapheme* pada suku kata tergantung dengan tipe huruf konsonan dan vokal apakah termasuk horizontal atau vertikal. Bentuk penulisan secara dua arah membuat Hangul memiliki tingkat kompleksitas yang lebih tinggi dibandingkan dengan karakter latin yang ditulis hanya dengan satu arah (Kim, Lee, dan Kim, 1993).

Kompleksitas karakter Hangul yang tinggi membuat banyak peneliti tertarik melakukan penelitian mengenai pengenalan karakter Hangul. Penelitian mengenai karakter Hangul dilakukan oleh Kim dan Jeong (1995) dengan dua jenis metode ekstraksi ciri, yaitu *directional segment feature* (DSF) dan *directional contributivity*

feature (DCF) serta metode *distance matching* dan model jaringan syaraf tiruan (JST) sebagai pengklasifikasi. Pemilihan metode *distance matching* dikarenakan metode ini sederhana dan efektif untuk memilih beberapa kandidat dari jumlah data yang besar, serta JST yang memiliki keunggulan dalam hal klasifikasi data dalam jumlah yang sedikit. Hasil akurasi yang didapat dalam penelitian ini adalah sebesar 84.1%.

Penggunaan model jaringan syaraf tiruan untuk mengenali karakter Hangul lainnya dilakukan oleh Trinanda (2013) dengan menggunakan ekstraksi ciri *Diagonal* dan *Backpropagation*. Menurut Trinanda, hasil pengujian menggunakan data latih mendapatkan keakuratan sebesar 95,21%, sedangkan hasil pengujian menggunakan data uji sebesar 78,96%. Kedua hasil tersebut menunjukkan bahwa karakter Hangul dapat dikenali baik menggunakan data latih maupun data uji, meskipun untuk data uji, hasilnya belum memuaskan.

Metode *Zone Centroid and Zone (ZCZ)* telah sering digunakan sebagai ekstraksi ciri dalam pengenalan karakter tulisan tangan. Studi kasus mengenai penggunaan Metode ZCZ dilakukan oleh Rajashekararadhya dan Ranjan (2008) dalam mengenali empat *script* populer India Selatan, yaitu Telugu, Kannada, Malayalam, dan Tamil. Dalam penelitian tersebut, ZCZ dibandingkan dengan metode *Image Centroid and Zone (ICZ)* dan *Hybrid ZCZ + ICZ*. Hasil penelitian menggunakan ZCZ sebagai ekstraksi ciri menunjukkan akurasi yang tinggi dalam mengenali keempat *script* tersebut, yaitu sebesar 96% dalam mengenali digit Kannada, 94,5% dalam mengenali digit Telugu, 95% dalam mengenali digit Tamil, dan 94% dalam mengenali digit Malayalam.

Berdasarkan uraian di atas, penelitian ini akan mengembangkan sebuah perangkat lunak untuk mengenali karakter Hangul tulisan tangan dengan menggunakan metode *Zone Centroid and Zone (ZCZ)* sebagai metode ekstraksi ciri dan *Backpropagation* sebagai metode pengenalannya. Pengenalan karakter Hangul tulisan tangan menggunakan metode *Zone Centroid and Zone (ZCZ)* dan *Backpropagation* ini diharapkan mampu mencapai tingkat akurasi tinggi.

1.2 Rumusan Masalah

Pengenalan karakter tulisan tangan karakter Hangul merupakan hal yang tidak sederhana karena bentuk karakter Hangul yang unik dan berbeda dengan karakter latin pada umumnya. Karakter Hangul berbentuk simbol tersendiri dan bentuk penulisannya juga membuatnya berbeda dari karakter lainnya. Selain itu, setiap individu memiliki variasi dalam penulisan seperti gaya, ukuran, bentuk, dan ketebalannya. Bentuk penulisan secara dua arah membuat Hangul memiliki tingkat kompleksitas yang lebih tinggi dibandingkan dengan alfabet Latin yang ditulis hanya dengan satu arah.

1.3 Tujuan Penelitian

Tujuan penelitian ini adalah mengembangkan perangkat lunak untuk mengenali tulisan tangan karakter Hangul dengan menerapkan metode *Zone Centroid and Zone (ZCZ)* dan *Backpropagation* serta mengetahui tingkat akurasinya.

1.4 Manfaat Penelitian

Manfaat yang diperoleh dari penelitian ini adalah:

1. Hasil penelitian berupa perangkat lunak pengenalan karakter Hangul tulisan tangan yang dapat digunakan untuk membantu proses *Handwritten Character Recognition (HCR)*.
2. Hasil penelitian ini dapat dikembangkan lebih lanjut menjadi sistem pengenalan Hangul secara *online*.

1.5. Ruang Lingkup Masalah

Dalam penelitian ini, ditetapkan beberapa batasan masalah sebagai berikut :

1. Format citra masukan adalah JPEG (*.jpg)
2. Inputan berupa file *scanned* karakter Hangul tulisan tangan ditulis menggunakan tinta hitam.
3. Karakter Hangul yang dilatih terdiri dari 10 buah karakter dengan dua buah *grapheme*.
4. Parameter *Backpropagation* yang digunakan yaitu iterasi maksimal (*maximum iteration*) sebesar 100 juta, kecepatan belajar sebesar 0,4, dan *error tolerance* sebesar 0.001.
5. Variasi penulisan karakter Hangul yang diteliti adalah penulisan secara tegak lurus dan miring.
6. Proses penskalaan menggunakan *method drawImage()* dalam *library* bahasa pemrograman java.
7. Pengenalan karakter dilakukan secara *offline* berbasis *desktop*.

1.6. Metodologi Penelitian

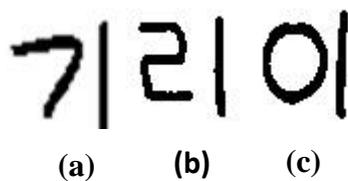
1.6.1. Unit Penelitian

Unit penelitian yang digunakan pada penelitian ini adalah Fakultas Ilmu Komputer Universitas Sriwijaya.

1.6.2. Metode Pengumpulan Data

1.6.2.1. Jenis Data

Jenis data yang digunakan dalam penelitian ini adalah data primer berupa citra tulisan tangan karakter Hangul. Data primer merupakan data yang diperoleh langsung dari subjek penelitian sebagai sumber informasi yang dicari. Citra masukan yang digunakan dalam format .JPG. Gambar I-1 menunjukkan contoh data set karakter Hangul tulisan tangan.



Gambar I-1 (a). Data Set Citra Karakter Hangul Tulisan Tangan “Gi”, (b). Data Set Citra Karakter Hangul Tulisan Tangan “Ri” (c). Data Set Citra Karakter Hangul Tulisan Tangan “i”

1.6.2.2. Teknik Pengumpulan Data

Sumber data yang digunakan untuk penelitian merupakan citra yang berasal dari tulisan tangan hangul pada kertas putih A4 menggunakan spidol hitam yang di-*scan* dengan *scanner* seperti pada gambar I-1. Data yang telah dikumpulkan akan digunakan pada penelitian ini sebagai data pelatihan (*training*) dan data uji (*testing*).

1.6.3. Metode Pengembangan Perangkat Lunak

Metode pembangunan perangkat lunak yang digunakan adalah *Rational Unified Process* (RUP) yang merupakan model pengembangan perangkat lunak berorientasi objek, dimana dengan metode ini dapat dilakukan perbaikan terhadap fase-fase sebelumnya. Terdapat empat fase dalam metode pembangunan perangkat lunak RUP, yakni:

1. Fase Insepsi

Pada fase insepsi hal-hal yang harus dilakukan adalah membuat *business modeling* dan *requirement* :

- *Business Modeling*:
 1. Memahami konsep dari pengenalan karakter Hangul tulisan tangan menggunakan metode *Zone Centroid and Zone* (ZCZ) dan *Backpropagation*.
 2. Menentukan kebutuhan perangkat lunak dan keras yang akan digunakan sebagai penunjang.
 3. Menggambarkan diagram *use case* tahap awal dan skenario dari perangkat lunak pengenalan karakter Hangul tulisan tangan menggunakan metode *Zone Centroid and Zone* (ZCZ) dan *Backpropagation*.
- *Requirement*:

Menentukan *requirement* pada perangkat lunak pengenalan karakter Korea tulisan tangan menggunakan metode *Zone Centroid and Zone* (ZCZ) dan *Backpropagation*. Kegiatan yang dilakukan yaitu mengumpulkan citra karakter Hangul tulisan tangan, melakukan proses prapengolahan citra,

dilanjutkan dengan mengekstraksi ciri citra karakter Hangul tulisan tangan, dan selanjutnya dilakukan pengenalan citra karakter Hangul tulisan tangan.

2. Fase Elaborasi

Hal – hal yang dilakukan pada fase elaborasi adalah :

1. Melakukan analisis dan perancangan perangkat lunak pengenalan karakter Hangul tulisan tangan menggunakan metode *Zone Centroid and Zone (ZCZ)* dan *Backpropagation*.
2. Menggambarkan diagram *use case* tahap awal.
3. Menggambarkan model kelas analisis, diagram kelas, dan *sequence diagram* dari perangkat lunak pengenalan karakter Hangul tulisan tangan menggunakan metode *Zone Centroid and Zone (ZCZ)* dan *Backpropagation*.

3. Fase Konstruksi

Tahapan yang akan dilakukan, yaitu :

1. Memastikan kelengkapan dan kesesuaian antara diagram *use case*, model kelas analisis, diagram kelas, dan *sequence diagram*.
2. Membuat kode program yang sesuai dengan fungsi-fungsi yang telah digambarkan pada fase sebelumnya.
3. Melakukan pengujian perangkat lunak dan perbaikan berdasarkan hasil analisis pengujian.

4. Fase Transisi

Pada fase ini akan dilakukan pengujian terhadap perangkat lunak dengan metodologi pengujian perangkat lunak yaitu, metode *black box testing*.

1.7. Sistematika Penulisan

Sistematika penulisan tugas akhir ini mengikuti standar penulisan tugas akhir Fakultas Ilmu Komputer Universitas Sriwijaya, sebagai berikut :

BAB I. PENDAHULUAN

Pada bab ini diuraikan mengenai latar belakang, perumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, metodologi penelitian, dan sistematika penulisan.

BAB II. LANDASAN TEORI

Pada bab ini akan dibahas dasar-dasar teori yang digunakan dalam melakukan analisis, perancangan, dan implementasi tugas akhir yang dilakukan pada bab-bab selanjutnya.

BAB III. ANALISIS DAN PERANCANGAN

Pada bab ini akan dibahas mengenai analisis serta perancangan terhadap algoritma yang digunakan dalam proses pengenalan tulisan tangan karakter Hangul dan perancangan tentang desain arsitektural dari perangkat lunak yang akan dibuat, sehingga dapat membantu dalam melakukan implementasi nantinya.

BAB IV. HASIL DAN PEMBAHASAN

Pada bab ini akan dibahas mengenai lingkungan implementasi perancangan dan analisis dari perangkat lunak yang akan dibuat, implementasi tujuan perangkat lunak, hasil eksekusi, dan hasil pengujian.

BAB V. KESIMPULAN DAN SARAN

Pada bab ini berisi kesimpulan dari semua uraian-uraian pada bab-bab sebelumnya dan juga berisi saran-saran yang diharapkan berguna dalam pengembangan perangkat lunak ini.