

1. Pendahuluan

1.1. Latar Belakang

Seiring dengan perkembangan teknologi dan informasi yang pesat diseluruh penjuru dunia, rasa keingintahuan masyarakat terhadap budaya dan bahasa dari negara lainpun ikut meningkat. Banyak negara yang dalam penulisan huruf bahasanya tidak menggunakan huruf romawi, melainkan menggunakan bentuk huruf lainnya. Adanya perbedaan penulisan huruf ini mempersulit masyarakat luar dalam proses pembelajaran budaya dan bahasa negara tersebut. Agar masyarakat luar dapat mempelajari budaya negara yang memiliki sistem penulisan huruf yang berbeda, maka masyarakat terpaksa memiliki dan memahami pengetahuan tata bahasa lokal dengan baik. Salah satu negara yang menggunakan sistem penulisan yang berbeda adalah negara Korea.

Negara Korea, terutama Korea Selatan, memiliki daya tarik yang kuat pada beberapa tahun terakhir. *Korean Wave* membuat masyarakat asing tertarik untuk mengenali budaya Korea lebih mendalam. Pengetahuan budaya Korea Selatan dapat ditemukan pada berbagai macam literatur, termasuk tulisan-tulisan yang terdapat pada media Internet. Tidak terlepas keharusan membaca literatur dalam bahasa Korea jika masyarakat ingin mengenali budaya Korea secara lebih mendalam. Oleh karenanya, masyarakat asing tentunya diharuskan memiliki kemampuan untuk dapat membaca dan memahami literatur yang tertulis dalam hangul guna mempelajari budaya Korea Selatan secara lebih mendalam.

Google Translate adalah sebuah alat bantu yang dapat mengartikan tulisan hangul Korea yang tertulis dalam media elektronik kedalam bahasa lain, seperti bahasa Indonesia ataupun bahasa Inggris. Kendala yang dihadapi adalah ketika suatu literatur yang ingin dibaca tidaklah tertulis pada media elektronik, melainkan pada media cetak, seperti buku, surat kabar, dan majalah.

Pengenalan Karakter Optik atau lebih dikenal dengan *Optical Character Recognition* (OCR) merupakan sebuah solusi yang dapat digunakan dalam mengenali karakter huruf dari sebuah gambar yang kemudian dikeluarkan kembali dalam bentuk teks[2]. Banyak penelitian OCR yang sudah dilakukan dalam pengenalan huruf Hangul, yang kemudian diterapkan menjadi sebuah aplikasi. Tingkat pengenalan rata-rata pada berbagai penelitian terkait sudah sangat tinggi. Meski demikian, masih terdapat kesalahan yang ditemui dalam pengenalan huruf Hangul Korea. Kesalahan pengenalan pada umumnya disebabkan oleh banyaknya susunan kombinasi huruf Hangul yang tersedia yang mengakibatkan

sulitnya proses segmentasi huruf. Jumlah suku kata yang dapat dibentuk dari kombinasi huruf Hangul mencapai 11172 karakter, namun hanya 2350 suku kata yang digunakan pada penulisan sehari-hari[7]. Metode yang umum digunakan pada pengenalan huruf Hangul adalah metode *template matching*[2][8]. Pada sebuah jurnal, dituliskan contoh kesalahan pengenalan huruf Hangul dengan menggunakan metode 6 bentuk *graphemes*. Pada jurnal tersebut, *recognition rate* pada bentuk *graphemes* ke-2 adalah 99.04%, yang dapat terbilang masih terhitung rendah dibandingkan dengan ke-5 *graphemes* lainnya yang mencapai 100%[7]. Hal ini terutama disebabkan oleh kesalahan segmentasi gambar yang menyebabkan potongan-potongan huruf tidak dapat dikenali dengan tepat. Hal inilah yang mendasari penulis untuk meneliti metode pengenalan huruf Hangul yang dapat menghasilkan tingkat pengenalan yang tinggi.

Random Forest diduga dapat meningkatkan akurasi pengenalan huruf Hangul. Pada metode ini, akan disiapkan sejumlah data latih yang selanjutnya akan digunakan sebagai *database* pada pencocokan karakter pada tahap pengenalan/pengujian suku kata.

1.2. Perumusan Masalah

Rumusan masalah yang akan diteliti pada Tugas Akhir ini adalah :

1. Bagaimana cara membuat sistem pengenalan karakter huruf Hangul Korea cetak dengan akurasi yang tinggi?
2. Metode apakah yang lebih cocok digunakan pada *random forest* untuk mendapatkan hasil pengenalan dengan akurasi tinggi?

1.3. Tujuan

Tujuan yang ingin dicapai dari Tugas Akhir ini antara lain :

1. Merancang sistem pengenalan karakter Hangul Korea dengan nilai akurasi yang tinggi.
2. Merancang sistem pengenalan karakter Hangul Korea dengan menggunakan berbagai jenis metode.

1.4. Batasan Masalah

Ruang Lingkup dan batasan masalah yang digunakan pada penelitian ini adalah:

1. Huruf Hangul yang digunakan berjumlah 2350 suku kata yang umum digunakan dari 11172 kemungkinan kombinasi suku kata sesuai dengan KS5602.
2. Huruf Hangul yang dianalisis dan diuji adalah teks cetak, sehingga tidak ada pengujian untuk tulisan tangan.
3. Posisi huruf Hangul yang digunakan adalah *front view*, sehingga tidak ada pengujian untuk bentuk transformasi.

4. Font huruf Hangul yang digunakan adalah Myunjo (Batang), Gothic (Dotum), dan Naru (Gulim).
5. Huruf Hangul yang dijadikan data latih adalah 3 bentuk font dengan masing-masing 2350 suku kata. Serta 3 bentuk font dengan masing-masing 2350 suku kata, yang telah diberi noise.
6. Latar belakang objek polos berwarna putih.

1.5. Metodologi Penelitian

Metodologi penelitian yang akan digunakan pada Tugas Akhir ini adalah sebagai berikut :

1. Studi Literatur

Studi literatur dilakukan dengan cara mengumpulkan dan mempelajari referensi yang berkaitan dengan Pengenalan Karakter Optik atau *Optic Character Recognition* (OCR) dengan fitur ekstraksi yang baik untuk pengenalan karakter Korea, melalui berbagai media, seperti Internet, paper, jurnal, dan buku.

2. Pengumpulan Data

Berbagai macam kombinasi karakter Hangul yang masuk dalam ruang lingkup pengujian Tugas Akhir ini akan dikumpulkan dan dibagi menjadi dua jenis data yaitu data latih dan data uji.

3. Implementasi Sistem

Implementasi sistem akan dibuat menggunakan MATLAB sesuai dengan perancangan sistem yang telah dibuat.

4. Analisis Hasil Pengujian

Hasil pengujian yang dianalisis pada penelitian ini adalah nilai *recognition rate* dan waktu komputasi.

5. Penyusunan Laporan Akhir

Laporan akhir disusun berdasarkan analisis seluruh hasil data latih dan data uji yang didapatkan selama proses penelitian ini.

1.6. Sistematika Penulisan

Buku Tugas Akhir ini dituliskan dengan sistematika agar pembahasan suatu masalah dapat diuraikan dengan jelas. Oleh karena itu, penulis membaginya kedalam 5 bab, yaitu :

BAB I PENDAHULUAN

Bab I berisi tentang latar belakang, rumusan masalah, tujuan,

batasan masalah, metodologi penelitian, dan sistematika penulisan.

BAB II LANDASAN TEORI

Bab II berisi tentang dasar teori bentuk-bentuk karakter Kanji yang dapat dikaitkan dengan Random Forest untuk diujikan pada Bab IV.

BAB III ANALISIS PERANCANGAN DAN IMPLEMENTASI

Bab III berisi tentang proses analisis perancangan alur kerja sistem yang kemudian diimplementasikan dan diujikan pada Bab IV.

BAB IV PENGUJIAN DAN ANALISIS

Bab IV berisi tentang tujuan dan analisis hasil pengujian sistem. Pengujian dibagi kedalam 3 skenario untuk mendapatkan tujuan yang telah ditentukan.

BAB V KESIMPULAN DAN SARAN

Bab V berisi tentang kesimpulan dari hasil analisis pengujian, serta saran yang berhubungan dengan pengembangan sistem lebih lanjut.