

**PENGENALAN POLA TULISAN HIRAGANA DENGAN
ADAPTIVE THRESHOLDING DAN PROJECTION
HISTOGRAM**

Skripsi



Disusun Oleh

ALDO KURNIA WICAKSONO

71160003

PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA
YOGYAKARTA

2020

PENGENALAN POLA TULISAN HIRAGANA DENGAN ADAPTIVE THRESHOLDING DAN PROJECTION HISTOGRAM

Skripsi



Diajukan kepada Program Studi Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana
Sebagai Salah Satu Syarat dalam Memperoleh Gelar
Sarjana Komputer

Disusun Oleh

ALDO KURNIA WICAKSONO

71160003

PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA YOGYAKARTA
2020

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI
SKRIPSI/TESIS/DISERTASI UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademika Universitas Kristen Duta Wacana, saya yang bertanda tangan di bawah ini:

Nama : Aldo Kurnia Wicaksono
NIM : 71160003
Program studi : Informatika
Fakultas : Teknologi Informasi
Jenis Karya : Skripsi

demikian pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Kristen Duta Wacana **Hak Bebas Royalti Noneksklusif** (*None-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul:

**“PENGENALAN POLA TULISAN HIRAGANA DENGAN ADAPTIVE
THRESHOLDING DAN PROJECTION HISTOGRAM”**

berserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti/Noneksklusif ini Universitas Kristen Duta Wacana berhak menyimpan, mengalih media/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama kami sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Yogyakarta
Pada Tanggal : 4 Juni 2020

Yang menyatakan



Aldo Kurnia Wicaksono
71160003

HALAMAN PENGESAHAN

SKRIPSI

Pengenalan Pola Tulisan Hiragana Dengan Adaptive Thresholding Dan Projection Histogram

Oleh: 71160003 / Aldo Kurnia Wicaksono

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 . . .

Yogyakarta, 7 Juli 2020

Mengesahkan

Dewan Penguji:

1. Aditya Wikan Mahastama, S.Kom., M.Cs.
2. Lucia Dwi Krisnawati, Dr. Phil.
3. Yuan Lukito, S.Kom., M.Cs.
4. Kristian Adi Nugraha, S.Kom., M.T.

Dekan

Ketua Program Studi



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

(Gloria Virginia, Ph.D.)

PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul:

Pengenalan Pola Tulisan Hiragana dengan Adaptive Thresholding dan Projection Histogram

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 kesarjanaan 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 kesarjanaan saya.

Yogyakarta, 4 Juni 2020



ALDO KURNIA WICAKSONO
71160003

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PERSETUJUAN.....	iv
HALAMAN PENGESAHAN.....	v
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xi
DAFTAR GAMBAR.....	xii
BAB 1 PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	1
1.3 Batasan Masalah.....	2
1.4 Tujuan Penelitian.....	2
1.5 Metodologi Penelitian.....	2
1.6 Sistematika Penulisan.....	3
BAB 2 TINJAUAN PUSTAKA DAN LANDASAN TEORI.....	4
2.1 Tinjauan Pustaka.....	4
2.2 Landasan Teori.....	6
2.2.1 Aksara Jepang.....	6
2.2.2 Huruf Hiragana.....	6
2.2.3 Optical Character Recognition.....	8
2.2.4 Pra-proses.....	8
2.2.5 Ekstraksi fitur.....	11
2.2.6 Metode Pengenalan.....	11
2.2.6.1 Penghitungan Jarak.....	11
2.2.6.2 Proses Klasifikasi.....	13
BAB 3 METODOLOGI PENELITIAN DAN PERANCANGAN SISTEM.....	15
3.1 Analisis Kebutuhan Program.....	15
3.1.1 Spesifikasi Perangkat Keras (<i>Hardware</i>).....	16
3.1.2 Spesifikasi Perangkat Lunak (<i>Software</i>).....	16
3.1.3 Kebutuhan <i>Library</i>	16

3.2 Blok Diagram Sistem	16
3.3 Rancangan Desain Antarmuka Program	18
3.4 Perancangan Pengujian Program.....	20
BAB 4 IMPLEMENTASI DAN ANALISIS PROGRAM	21
4.1 Implementasi Program	21
4.1.1 Tampilan Antarmuka	21
4.1.2 Implementasi Sistem	23
4.1.3 Proses Segmentasi.....	25
4.1.4 Lokal <i>Adaptive Thresholding</i>	26
4.1.5 Hasil Uji Program	27
4.2 Analisis Hasil Eksperimen	38
4.2.1 Persentase Pengenalan Karakter	38
4.2.2 <i>Projection Histogram, Euclidean Distance, dan K-NN</i>	39
BAB 5 KESIMPULAN DAN SARAN.....	43
5.1 Kesimpulan	43
5.2 Saran.....	44
DAFTAR PUSTAKA	45
LAMPIRAN.....	47

DAFTAR TABEL

Tabel 2.1 Huruf Hiragana dasar yang akan digunakan	7
Tabel 4.1 Tabel confusion matrix program dengan menggunakan $k = 9$ pada k-NN	28
Tabel 4.2 Hasil keseluruhan setelah pengujian program	38

©UKDW

DAFTAR GAMBAR

Gambar 2.1 Proses pra-proses.....	9
Gambar 2.2 Contoh adaptive thresholding.....	9
Gambar 2.3 Contoh adaptive thresholding.....	9
Gambar 2.4 Contoh adaptive thresholding.....	10
Gambar 2.5 Proses penilaian indeks vektor	11
Gambar 3.1 Blok Diagram Sistem	17
Gambar 3.2 Rancangan Desain Antarmuka Web	18
Gambar 3.3 Rancangan Desain Antarmuka Web Tentang Saya.....	18
Gambar 3.4 Rancangan Desain Antarmuka Web Proses	19
Gambar 3.5 Rancangan Desain Antarmuka Web Hasil	19
Gambar 4.1 Antarmuka Web Halaman Utama	21
Gambar 4.2 Antarmuka Web Tentang Saya	22
Gambar 4.3 Antarmuka Web Proses.....	22
Gambar 4.4 Antarmuka Web Hasil.....	23
Gambar 4.5 Proses Segmentasi Gagal	26
Gambar 4.6 Proses Trimming	26
Gambar 4.7 Tabel Confusion Matrix	27
Gambar 4.8 Tabel Confusion Matrix untuk karakter C	28
Gambar 4.9 Grafik akurasi pengenalan karakter.....	38
Gambar 4.10 Posisi dan panjang goresan dalam karakter menjadi parameter penentu dalam penghitungan jarak.....	39
Gambar 4.11 Proses penghitungan jarak karakter uji a dengan karakter latih he ..	41

INTISARI

Pengenalan Pola Tulisan Hiragana Dengan Adaptive Thresholding Dan Projection Histogram

Aksara Jepang sendiri terbagi menjadi tiga, yaitu aksara Kanji, aksara Hiragana, dan aksara Katakana. Biasanya, di negara Jepang sendiri aksara Katakana dan aksara Hiragana sudah dipelajari semenjak berada pada taman kanak-kanak atau disebut TK.

Bagi orang Indonesia tentu saja tidak mudah mempelajari aksara Jepang dikarenakan orang Indonesia lebih terbiasa menulis huruf dalam bentuk alfabet, sehingga orang Indonesia lebih mudah mempelajari aksara Inggris daripada aksara Jepang. Hal tersebut mengakibatkan orang Indonesia memerlukan waktu lebih lama dalam mempelajari serta menghafal aksara Hiragana.

Pada penelitian ini, hasil terbaik pada program pengenalan karakter yang dibuat adalah akurasi 43,91%, *error rate* 56,09%, dan presisi 51,40% dengan jumlah k pada k -NN yang digunakan yaitu 9. Untuk hasil kurang baik pada pengenalan karakter adalah akurasi 41,96%, *error rate* 58,04%, dan presisi 50,86% dengan jumlah k pada k -NN yang digunakan yaitu 10.

Kata Kunci : Pengenalan karakter, Aksara Jepang, Hiragana, *adaptive thresholding*, *projection histogram*, *euclidean distance*, *k*-NN.

ABSTRACT

Recognition Of Hiragana Writing Patterns With Adaptive Thresholding And Projection Histogram

Japanese characters themselves are divided into three, namely the Kanji script, the Hiragana script, and the Katakana script. Usually, in Japan itself the Katakana and Hiragana characters have been studied since they were in kindergarten or called kindergarten.

For Indonesians, of course it is not easy to learn Japanese characters because Indonesians are more accustomed to writing letters in the form of the alphabet, so that Indonesian people are easier to learn English characters than Japanese characters. This resulted in Indonesians taking longer to learn and memorize the Hiragana script.

In this study, the best results on the character recognition program made were 43.91% accuracy, 56.09% error rate, and 51.40% precision with the number of k in the k -NN used, which is 9. For poor results on recognition the characters are accuracy 41.96%, error rate 58.04%, and precision 50.86% with the number of k on the k -NN used is 10.

Keywords : Character Recognition, Japanese Characters, Hiragana, *adaptive thresholding, projection histogram, euclidean distance, k-NN.*

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Aksara Jepang sendiri terbagi menjadi tiga, yaitu aksara Kanji, aksara Hiragana, dan aksara Katakana. Biasanya, di negara Jepang sendiri aksara Katakana dan aksara Hiragana sudah dipelajari semenjak berada pada taman kanak-kanak atau disebut TK. Aksara Jepang semakin populer dengan adanya *manga* dan *anime*. Namun, aksara Jepang sendiri cukup rumit untuk dipelajari bagi masyarakat Indonesia. Hal ini bisa diketahui dari pola karakter aksara yang unik yang berbeda dari huruf yang digunakan dalam bahasa Indonesia (Mawitjere, 2017).

Bagi orang Indonesia tentu saja tidak mudah mempelajari aksara Jepang dikarenakan orang Indonesia lebih terbiasa menulis huruf dalam bentuk alfabet, sehingga orang Indonesia lebih mudah mempelajari alfabet atau huruf Latin daripada aksara yang digunakan dalam bahasa Jepang. Hal tersebut mengakibatkan orang Indonesia memerlukan waktu lebih lama dalam mempelajari serta menghafal huruf Hiragana (Mawitjere, 2017).

Dengan seiring berkembangnya teknologi, masalah tersebut dapat diatasi dengan dibuatnya program pengenalan tulisan huruf Jepang khususnya huruf Hiragana menggunakan *adaptive thresholding*. Program ini bertujuan untuk mengenali tulisan tangan huruf *Hiragana* yang kemudian diolah dalam bentuk citra digital.

Oleh karena itu, program ini diharapkan dapat membantu orang untuk dapat mempelajari huruf Jepang khususnya Hiragana dengan lebih mudah melalui *input* tulisan tangan yang kemudian diolah dalam bentuk citra digital.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang permasalahan, secara garis besar, masalah tersebut dapat diuraikan sebagai berikut :

1. Bagaimana menerapkan *adaptive thresholding* dengan fitur *projection histogram* dalam keseluruhan proses pengenalan huruf Hiragana ?
2. Bagaimana tingkat akurasi *euclidean distance* dan *k-NN* yang diterapkan bersama *adaptive thresholding* dengan fitur *projection histogram* dalam pengenalan huruf Hiragana ?

1.3 Batasan Masalah

Berdasarkan rumusan masalah, batasan masalah yang akan dibahas adalah sebagai berikut :

1. Program yang dibuat adalah program pengenalan pola tulisan tangan huruf Hiragana.
2. Huruf Hiragana yang akan dikenali berjumlah 46 huruf dasar, mulai dari “a” sampai “n”.
3. Input berupa tulisan tangan huruf Hiragana dasar.
4. Input merupakan hasil scan dengan posisi horizontal dan tidak miring.
5. Tipe *derau* yang akan ditangani berupa *derau* ringan dan tidak mencakup *derau* tinta tembus atau jamur.

1.4 Tujuan Penelitian

Tujuan penelitian ini adalah membangun sistem Pengnalan Karakter Optik (Optical Character Recognition – OCR) yang menerapkan *adaptive thresholding* untuk pengenalan serta mengukur kinerja *adaptive thresholding* dalam proses aliran kerja OCR.

1.5 Metodologi Penelitian

OCR mengambil gambar teks sebagai input dan menghasilkan citra digital. Sistem OCR menggunakan langkah: pra-proses (*thresholding*), ekstraksi fitur, dan klasifikasi. Dengan menggunakan 10 data latih, dan 1 data uji pada tiap huruf Hiragana. Pra-proses dan ekstraksi fitur, dan klasifikasi dilakukan dalam kedua data. Gambar teks disegmentasi menggunakan teknik *adaptive thresholding* yang

kemudian diekstraksi dengan fitur *projection histogram* lalu menjalankan proses klasifikasi menggunakan *euclidean distance* dan k-NN dengan cara data uji dibandingkan dengan data latih untuk mendapatkan output yang diinginkan.

1.6 Sistematika Penulisan

Laporan penelitian ini terdiri dari 5 bab, yaitu : Pendahuluan, Tinjauan Pustaka, Metodologi Perancangan, Pembahasan, dan Kesimpulan dan Saran. Penjelasan setiap bagian adalah sebagai berikut :

1. Pendahuluan

Bab ini terdiri dari Latar Belakang Masalah, Rumusan Masalah, Batasan Masalah, Tujuan Penulisan, Metodologi Penelitian, dan Sistematika Penulisan.

2. Tinjauan Pustaka

Bab ini berisi teori-teori yang berkaitan dengan pengenalan huruf *hiraagana* menggunakan *adaptive thresholding* fitur *projection histogram*.

3. Metodologi Perancangan

Bab ini berisi metode-metode dan teori yang akan diimplementasikan terkait perancangan sistem.

4. Pembahasan

Bab ini berisi pembahasan-pembahasan dari rancangan sistem yang telah dibuat pada Bab 3.

5. Kesimpulan dan Saran

Bab ini berisi kesimpulan akhir dari keseluruhan program dan saran-saran untuk pengembangan program

BAB 5

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Dari hasil penelitian yang telah dilakukan oleh peneliti, maka dapat disimpulkan bahwa *adaptive thresholding* dan *projection histogram* dapat diterapkan dengan cara pada *adaptive thresholding* melakukan *grayscale* citra terlebih dahulu, kemudian melakukan segmentasi untuk memotong tiap karakter pada citra, lalu karakter yang sudah tersegmentasi tersebut dihitung nilai thresholdnya dan diubah menjadi citra biner 1 jika diatas nilai threshold dan 0 jika dibawah nilai threshold. Pada *projection histogram*, melakukan penghitungan nilai biner 1 pada tiap baris dan kolom. Penerapan metode *adaptive thresholding*, *projection histogram* dengan *euclidean distance* dan k-NN dapat memberikan hasil pada k-NN sebanyak 9 dengan nilai akurasi 43.91%, *error rate* 56,09%, dan *presisi* 51,40% dari 460 data uji.

Hal yang menjadi kekurangan pada penelitian ini adalah sebagai berikut :

1. Ukuran, posisi dan goresan elemen karakter dalam tulisan tangan responden sangat bervariasi sehingga mempengaruhi jumlah piksel pada area tertentu. Ini mengakibatkan nilai jarak kakarter uji dengan karakter latih dari kelas yang sama kurang optimal.
2. Pemrosesan program memakan waktu yang lama dalam mengenali karakter. Hal ini dikarenakan oleh besarnya ukuran piksel gambar, sehingga program harus membaca tiap baris piksel agar bisa melanjutkan ke langkah berikutnya.
3. Ukuran tulisan responden tidak dibatasi sehingga menyebabkan ukuran karakter yang berbeda-beda (seperti ada yang ukuran sedang, besar, dan terlalu besar), sehingga pada proses *resize* terdapat piksel yang hilang dan berpengaruh pada bentuk karakter tersebut.

Kekurangan dan masalah pada penelitian ini diharapkan dapat disempurnakan oleh penelitian selanjutnya

5.2 Saran

Program yang telah dibuat masih dapat dikembangkan lagi dari banyak sisi yang dapat membuat program menjadi lebih baik. Beberapa hal tersebut dan kekurangan yang ada, penulis memberikan beberapa saran yang dapat digunakan untuk mengembangkan program lebih bagus lagi dan menaikkan akurasi dari penelitian ini, yaitu :

1. Jumlah data latih diperbanyak agar pengenalan karakter lebih baik.
2. Dalam proses uji coba, variasi nilai k pada k -NN disarankan menggunakan nilai $k = 7-9$. Semakin tinggi nilai k , semakin bagus hasilnya karena memiliki data yang cukup untuk *voting*. Tetapi jika terlalu tinggi, hasil menjadi kurang bagus.
3. Dalam penelitian ini, gambar diukur ulang dengan dimensi 75×75 piksel. Berdasarkan eksperimen dan hasil, ukuran ini terlalu besar sehingga waktu proses menjadi lama. Pada penelitian selanjutnya disarankan untuk menggunakan ukuran piksel yang lebih kecil dan lebih terukur.
4. Menetapkan batasan ukuran tulisan tangan responden. Hal ini diharapkan dapat menaikkan akurasi dalam pengenalan karakter.

DAFTAR PUSTAKA

- Andono, P. N., Sutojo, T., & Muljo no. (2017). *Pengolahan Citra Digital*. Yogyakarta: Penerbit ANDI.
- Arifianto, T. (2017). Segmentasi Aksara Pada Tulisan Aksara Jawa. *SMATIKA Jurnal*, 1-5.
- Cahyani, S., Wiryasaputra, R., & Gustriansyah, R. (2018). Identifikasi Huruf Kapital Tulisan Tangan Menggunakan Linear Discriminant Analysis dan Euclidean Distance. *Jurnal Sistem Informasi Bisnis*, 57-67.
- Divakaran, D. (2012). Spectral Analysis of Projection Histogram for Enhancing Close matching character Recognition in Malayalam. *International Journal of Computer Science & Information Technology (IJCSIT)*, 141-147.
- Istiqomah, D., Diner, L., & Wardhana, C. K. (2015). Analisis Kesulitan Belajar Bahasa Jepang Siswa Smk Bagimu Negeriku Semarang. *Journal of Japanese Learning and Teaching*, 1-4.
- Jana, R., Chowdhury, C., & Islam, M. (2013). Optical Character Recognition from Text Image. *International Journal of Computer Applications Technology and Research*, 239 - 243.
- Krisnawati, L. D., & Mahastama, A. W. (2019). Building Classifier Models for on-off Javanese Character Recognition. *iiWAS2019*, 1-2.
- Mahto, M. K., Bhatia, K., & Sharma, R. K. (2015). Combined Horizontal and Vertical Projection Feature Extraction Technique for Gurmukhi Handwritten Character Recognition. *International Conference on Advances in Computer Engineering and Applications (ICACEA)*, 1-8.

- Mawitjere, I. (2017). Peningkatan Penguasaan Huruf Jepang (Hiragana Dan Katakana) Melalui Media Buku Bergambar. *KOMPETENSI Jurnal Ilmiah Bahasa dan Seni*, 1-10.
- Naser, M., Hamid, N. I., & Hoque, M. (2014). Projection Based Feature Extraction Process For Bangla Script: A Modified Approach. *Software Technology and Engineering*, 210-214.
- Roy, Payel; et al. (2014). Adaptive Thresholding: A Comparative Study. *International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT)*, 1320-1324.
- Sutramiani, N. P., Putra, I. K., & Sudarma, M. (2015). Local Adaptive Thresholding Pada Pra-proses Citra Lontar Aksara Bali. *JURNAL TEKNOLOGI ELEKTRO*, 1693-2951.