

ARTIKEL

**PENGENALAN POLA TULISAN TANGAN HURUF JEPANG
(HIRAGANA) MENGGUNAKAN METODE *K-NEAREST NEIGHBOUR***



Oleh:

ARISKA FITRIA ANGGELINA

14.1.03.02.0294

Dibimbing oleh :

- 1. Ahmad Bagus Setiawan, S.T., M.Kom., M.M.**
- 2. Ardi Sanjaya, M.Kom.**

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS NUSANTARA PGRI KEDIRI

TAHUN 2018

SURAT PERNYATAAN
ARTIKEL SKRIPSI TAHUN 2018

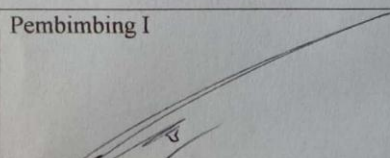
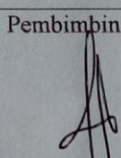
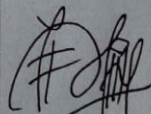
Yang bertanda tangan di bawah ini:

Nama Lengkap : Ariska Fitria Anggelina
NPM : 14.1.03.02.0294
Telepon/HP : 085736324094
Alamat Surel (Email) : riska1896@gmail.com
Judul Artikel : Pengenalan Pola Tulisan Tangan Huruf Jepang
(Hiragana) Menggunakan Metode *K-Nearest Neighbour*
Fakultas – Program Studi : Teknik – Teknik Informatika
Nama Perguruan Tinggi : Universitas Nusantara PGRI Kediri
Alamat Perguruan Tinggi : JL. K.H. Achmad Dahlan No.76 Telepon (0354) 771503
Kota Kediri.

Dengan ini menyatakan bahwa :

- artikel yang saya tulis merupakan karya saya pribadi (bersama tim penulis) dan bebas plagiarisme.
- artikel telah diteliti dan disetujui untuk diterbitkan oleh Dosen Pembimbing I dan II.

Demikian surat pernyataan ini saya buat dengan sesungguhnya. Apabila di kemudian hari ditemukan ketidaksesuaian data dengan pernyataan ini dan atau ada tuntutan dari pihak lain, saya bersedia bertanggungjawab dan diproses sesuai dengan ketentuan yang berlaku.

Mengetahui		Kediri, 31 Januari 2018
Pembimbing I  Ahmad Bagus Setiawan, S.T., M.Kom., M.M. NIDN. 0703018704	Pembimbing II  Ardi Sanjaya, M.Kom. NIDN. 0706118101	Penulis,  Ariska Fitria A. NPM. 14.1.03.02.0286

Ariska Fitria Anggelina | 14.1.03.02.0294
Fakultas Teknik – Teknik Informatikasimki.unpkediri.ac.id
|| 2 ||

PENGENALAN POLA TULISAN TANGAN HURUF JEPANG (HIRAGANA) MENGGUNAKAN METODE *K-NEAREST NEIGHBOUR*

Ariska Fitria Angelina

14.1.03.02.0294

Fakultas Teknik – Teknik Informatika

riska1896@gmail.com

Ahmad Bagus Setiawan, S.T., M.Kom., M.M. dan Ardi Sanjaya, M.Kom.

UNIVERSITAS NUSANTARA PGRI KEDIRI

ABSTRAK

Penelitian ini dilatar belakangi dari pengamatan dan pengalaman peneliti, bahwa bentuk tulisan tangan setiap orang tidak sama. Tulisan tangan seseorang sering berubah-ubah setiap waktu karena dipengaruhi beberapa faktor, misalnya waktu, umur dan kebiasaan. Beberapa orang memiliki kemampuan dalam mengenali tulisan tangan seseorang melalui pembelajaran. Namun berbeda jika yang mengenali tulisan tersebut adalah komputer. Mungkin ada beberapa kendala yang ditemukan saat pengenalan, sehingga tulisan tangan seseorang menjadi sulit dikenali.

Permasalahan penelitian ini adalah (1) Bagaimana merancang dan membangun suatu sistem yang dapat digunakan sebagai pengenalan perkembangan pola tulisan tangan huruf *Hiragana* dasar (*seion*) ?. (2) Bagaimana menerapkan metode Partisi Citra untuk melakukan ekstraksi fitur pada citra digital ?. (3) Bagaimana mengimplementasikan metode *K-Nearest Neighbour* dengan perhitungan jarak *euclidean distance* untuk mengidentifikasi data huruf *Hiragana* dasar (*seion*) ?.

Penelitian ini menggunakan metode Partisi sebagai pengekstraksi fitur citra, serta metode *K-Nearest Neighbour* sebagai klasifikasi dan perhitungan jaraknya menggunakan *Euclidean Distance*. Sebelum dilakukan pengenalan, *file* citra dilakukan proses *preprocessing* terlebih dahulu yaitu *grayscale* dan deteksi tepi *prewitt*, selanjutnya dilakukan partisi, lalu diidentifikasi.

Dari hasil pengujian pada skenario pertama dengan data *training* 60 dan *testing* mencakup semua tulisan minggu pertama sebanyak 90 huruf diperoleh akurasi sebesar 6%. Pada skenario kedua dengan jumlah data *training* 60 dan *testing* mencakup semua tulisan minggu kedua diperoleh akurasi sebesar 81%. Pada skenario ketiga dengan data *training* yang sama dan *testing* mencakup semua tulisan minggu ketiga diperoleh akurasi sebesar 88%. Dari hasil pengujian, dapat disimpulkan bahwa metode *K-Nearest Neighbour* dan Partisi dapat digunakan untuk mengidentifikasi tulisan tangan huruf *Hiragana*.

Kata kunci : Pengenalan Pola, Tulisan Tangan, Huruf *Hiragana*, *K-Nearest Neighbour*.

I. LATAR BELAKANG

Pada era modern saat ini menguasai bahasa asing merupakan salah satu kebutuhan yang harus dimiliki setiap orang. Banyak orang

yang berkunjung ke negara lain untuk bekerja, menempuh pendidikan atau bahkan hanya untuk sekedar menikmati liburan. Salah satu negara

yang banyak dikunjungi adalah negara Jepang (Nana Febriana dkk, tanpa tahun).

Kebudayaan Jepang sendiri saat ini banyak sekali diminati masyarakat Indonesia, beberapa diantaranya adalah *manga*, *anime*, film, musik, *fashion*, *make up* dan lain-lain. Bahasa Jepang merupakan salah satu bahasa yang lumayan sulit untuk dipelajari, karena kita harus mengerti keseluruhan hurufnya, tata cara penulisannya dan tata cara bacanya.

Negara Jepang memiliki bentuk huruf yang terdiri dari tiga kelompok yaitu *Katakana*, *Hiragana* dan *Kanji*. Huruf *Hiragana* dan *Katakana* dibuat oleh orang Jepang, sedangkan huruf *Kanji* berasal dari Cina yang kemudian mengalami perubahan cara baca dan cara penulisannya (Lalujan, tanpa tahun).

Karakter huruf Jepang merupakan karakter yang cukup kompleks dan memiliki karakteristik unik dibandingkan dengan karakter latin apabila ditulis dengan tulisan tangan, dimana terdapat keterkaitan antar piksel. Secara tradisional tulisan Jepang ditulis secara vertikal, dimulai dari sudut kanan atas kertas atau dengan cara lain yaitu horisontal dari kiri atas kertas. Untuk menulis huruf

Jepang dapat menggunakan aturan kanji dan *syllabaries* (Kana).

Pada dasarnya, tulisan tangan setiap orang memiliki sifat identik namun tidak sama. Artinya tulisan tangan seseorang sering berubah-ubah setiap waktu. Pada kenyataannya, perubahan-perubahan tersebut dipengaruhi oleh beberapa faktor, misalnya waktu, umur dan kebiasaan. Meski demikian, tulisan tangan seseorang dapat ditangani sebagai sebuah citra sehingga dapat dikenali dengan menggunakan aplikasi pengenalan pola dan pengolahan citra.

Beberapa peneliti telah menggunakan algoritma *K-Nearest Neighbour* untuk mengenali huruf dan angka. (Roni Akbar, Eko Adi Sarwoko, 2016) telah menggunakan metode *K-Nearest Neighbour* untuk mengenali pola tulisan tangan angka *Arabic* (Indian) untuk angka 0-9 diperoleh tingkat akurasi hingga 86%. Penelitian lainnya oleh (Zaky Mukhoyyar, tanpa tahun) dengan algoritma yang sama yaitu *K-Nearest Neighbour* untuk mengenali kata aksara Jawa dengan tingkat keberhasilan pengenalan 100% dan implementasi segmentasi berbasis histogram memiliki tingkat keberhasilan 90%. Sementara itu,

huruf yang sama yaitu huruf Jepang *hiragana* juga dijadikan objek pada skripsi dengan judul pengenalan pola huruf Jepang *hiragana* menggunakan algoritma *backpropagation* oleh (Nicolaus Euclides Wahyu Nugroho, 2016) dengan tingkat akurasi sebesar 86,63%.

Melihat tingkat keberhasilan yang cukup tinggi yang dihasilkan oleh para peneliti sebelumnya serta belum adanya sistem dengan algoritma *K-Nearest Neighbour* untuk mengenali pola tulisan tangan huruf Jepang *Hiragana* maka penulis mencoba menggunakan algoritma *K-Nearest Neighbour* untuk menguji dalam sistem pengenalan pola tulisan tangan huruf Jepang *Hiragana* dasar (*seion*).

II. METODE

A. Deteksi Tepi Prewitt

“Operator *Prewitt* merupakan pengembangan operator *Robert* dengan menggunakan *High Pass Filter* (HPF) yang diberi satu angka nol penyangga, serta menggunakan persamaan yang sama dengan operator *sobel* yang menggunakan matrik 3×3 ” (Gonzales, 2002). Operator ini mengambil prinsip dari fungsi *laplacian* yang dikenal sebagai fungsi untuk membangkitkan

HPF. Untuk mempercepat komputasi, bagian yang bernilai nol tidak perlu diproses.

B. Euclidean Distance

Menurut Darma Putra (2010) *euclidean distance* dimaknai sebagai berikut :

Euclidean Distance adalah matriks yang paling sering digunakan untuk menghitung kesamaan 2 vektor. Jarak *euclidean* menghitung akar dari kuadrat perbedaan 2 vektor (*root of square differences between 2 vector*).

Persamaan dari jarak *euclidean* adalah :

$$d_{ij} = \sqrt{\sum_{k=1}^n (x_{ik} - x_{jk})^2} \dots (1)$$

Hasil perhitungan *euclidean distance* ini akan memperlihatkan seberapa besar tingkat kesamaan antara citra uji dan citra sampel, semakin kecil nilai dari *euclidean distance* (mendekati nilai nol), maka tingkat kemiripan (*similarity*) citra semakin baik.

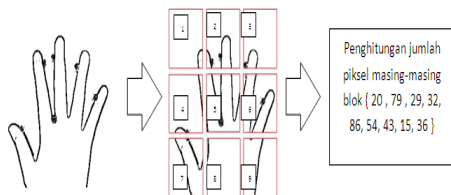
C. K-Nearest Neighbour

Ide dasar dari klasifikasi KNN (*K-Nearest Neighbor*) cukup sederhana, yaitu klasifikasi yang didasarkan pada kelas dari tetangga terdekat. Klasifikasi ini memungkinkan untuk mengambil lebih dari satu tetangga, teknik ini biasa disebut klasifikasi *K-Nearest Neighbor* dimana K

tetangga terdekat digunakan untuk menentukan kelas (cunningham dkk., 2007).

D. Partisi Citra

Partisi citra adalah membagi citra menjadi beberapa blok dimana masing-masing blok atau bagian memiliki ukuran yang sama besar. Tujuan dilakukan partisi terhadap citra adalah menghitung jumlah piksel masing-masing blok pada data *training* dan data *testing* (Ardi Sanjaya dkk., 2014).

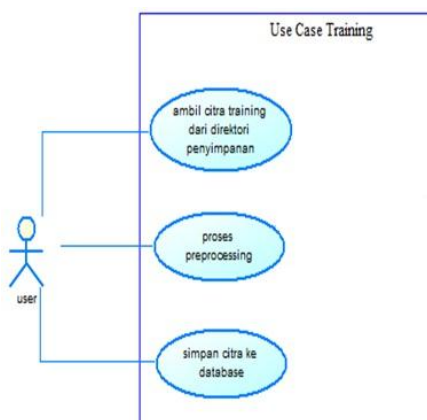


Gambar 1. Contoh Partisi Citra

III. HASIL DAN KESIMPULAN

A. Implementasi Sistem

1. Use Case Training



Gambar 1 : Use Case Training

Keterangan *use case training* pada gambar 2 adalah sebagai berikut :

a. Pengambilan citra *training*.

User menginputkan data citra *training* huruf *Hiragana* .

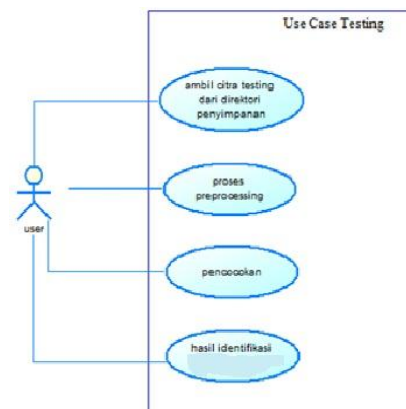
b. *Preprocessing* : *user*

melakukan *preprocessing* pada citra *hiragana*, dimana di dalam proses ini terdapat subproses yaitu *grayscale* dan deteksi tepi dengan operator *prewitt*, dilanjutkan dengan proses partisi citra untuk mendapatkan ciri dari masing-masing objek.

c. *Database* : digunakan untuk

menyimpan presentase piksel citra *training* ke dalam *database* yang telah melalui tahap *preprocessing* sebelumnya untuk dijadikan acuan data *testing*.

2. Use Case Testing



Gambar 2 : Use Case Testing

Keterangan *use case testing* pada gambar 2 adalah sebagai berikut :

- a. Pengambilan citra *training*.
User menginputkan data citra *training* huruf *Hiragana* .
- b. *Preprocessing* : *user* melakukan *preprocessing* pada citra *hiragana*, dimana di dalam proses ini terdapat subproses yaitu *grayscale* dan deteksi tepi dengan operator *prewitt*, dilanjutkan dengan proses partisi citra untuk mendapatkan ciri dari masing-masing objek.
- c. Pencocokan : proses pencocokan atau identifikasi menggunakan metode KNN dengan perhitungan jarak *Euclidean distance*.
- d. Hasil identifikasi : setelah dilakukan pencocokan data kemudian sistem akan menampilkan hasil berupa nama huruf yang dikenali beserta jarak atau nilai kedekatannya.

B. Tampilan Program

1. Tampilan *Home*
Home adalah *form* yang pertama kali muncul saat

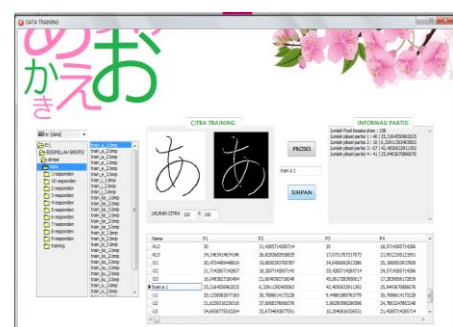
program dijalankan. *Form home* dapat dilihat pada gambar berikut :



Gambar 3 : Tampilan *Form Home*

2. Tampilan *Training*

Training adalah *form* yang dapat diakses dari *form Home*. *Form* ini digunakan sebagai akses untuk memproses data *training*. *Form training* dapat dilihat pada gambar berikut :



Gambar 4 : Tampilan *Form Training*

3. Tampilan Identifikasi

Identifikasi adalah *form* yang dapat diakses dari *form* beranda. *Form* ini digunakan sebagai

proses identifikasi. Tampilan *form* identifikasi dapat dilihat pada gambar berikut :



Gambar 5 : Tampilan *Form*
Identifikasi

C. Skenario Uji Coba

1. Skenario Uji Coba 1

Skenario ini menggunakan data *training* sebanyak 30 citra huruf *hiragana* dan data *testing* sebanyak 90, dengan rincian mengambil data minggu ke satu dari masing-masing responden.

2. Skenario Uji Coba 2

Skenario ini menggunakan data *training* sebanyak 30 citra huruf *hiragana* dan data *testing* sebanyak 90, dengan rincian mengambil data minggu ke dua dari masing-masing responden.

3. Skenario Uji Coba 3

Skenario ini menggunakan data *training* sebanyak 30 citra huruf *hiragana* dan data *testing* sebanyak 90, dengan rincian mengambil data

minggu ke tiga dari masing-masing responden.

4. Skenario Uji Coba 4

Skenario ini menggunakan data *training* sebanyak 30 citra *hiragana*, data *testing* berupa objek lain selain huruf *hiragana* dan menggunakan huruf *hiragana* diluar data yang diuji sebelumnya yaitu A, I, U, E, O, KA, KI, KU, KE, KO.

Dari hasil skenario uji coba diatas diperoleh kesimpulan sebagai berikut :

1. Skenario 1, dari data citra *testing* minggu pertama sebanyak 90 yang diuji, 52 dikenali dengan benar dan 38 lainnya bernilai salah. Sehingga didapatkan akurasi sebesar 57,78%.
2. Skenario 2, dari data citra *testing* minggu kedua sebanyak 90 yang diuji, 69 dikenali dengan benar dan 21 lainnya bernilai salah. Sehingga didapatkan akurasi sebesar 76,67%.
3. Skenario 3, dari data citra *testing* minggu pertama sebanyak 90 yang diuji, 74 dikenali dengan benar dan

16 lainnya bernilai salah. Sehingga didapatkan akurasi sebesar 82,22%.

4. Skenario 4, objek selain huruf *Hiragana* ada 3 yang dapat dikenali dan 2 tidak dapat diidentifikasi oleh sistem. Sementara objek citra huruf *Hiragana* diluar data huruf yang diuji tetap dapat dikenali, namun tentu saja pengenalannya bernilai salah.

Tabel keakuarasian hasil uji coba skenario diatas dapat dilihat pada tabel berikut :

Tabel 1. Akurasi Uji Coba

No	Skenario	Benar	Salah	Akurasi
1	1	52	38	57,78%
2	2	69	21	76,67%
3	3	74	16	82,22%
4	4	0	0	0
Rata-rata				72,22%

D. Kesimpulan

Dari hasil penelitian, perancangan, pembuatan dan pengujian aplikasi Pengenalan Pola Tulisan Tangan Huruf Jepang (*Hiragana*) Menggunakan Metode *K-Nearest Neighbour* dapat disimpulkan sebagai berikut :

1. Metode *K-Nearest Neighbour* dan partisi citra dapat

digunakan untuk identifikasi citra tulisan tangan huruf *Hiragana*.

2. Huruf yang dapat dikenali 100% adalah A, I, U, E, KI. Hal ini dibuktikan pada skenario 1, 2 dan 3, dimana semua huruf tersebut dapat dikenali dengan benar tanpa ada kesalahan satu pun.
3. Huruf yang paling banyak salah atau tidak dapat dikenali adalah huruf KO. Hal ini dibuktikan dari hasil uji coba skenario 1, 2 dan 3, dimana huruf KO hanya dapat dikenali dengan benar sebanyak 3 kali.
4. Rata-rata akurasi dari semua pengujian adalah 72,22%.

IV. DAFTAR PUSTAKA

- Achmad, Balza.2005, Teknik Pengolahan Cita Digital menggunakan Delphi. Yogyakarta: Ardi Publishing.
- Adimihardja, Mulyana. 2011. *Yasashii Nihongo: Bahasa Jepang untk Kelas XI SMA/MA Program Pilihan* (F.M. Erly, Eds.). Bandung: Grafindo Media Pratama.
- Akbar, R. dan Sarwoko, A.E. 2016. Studi Analisis Pengenalan Pola Tulisan Tangan Angka Arabic (Indian) Menggunakan Metode *K-Nearest Neighbors* dan *Connected Component Labeling*. *Jurnal Dinamika Rekayasa*, 12 (2): 45-51.

- Cahyo, Septian Dwi. 2009. *Analisis Perbandingan Beberapa Metode Deteksi Tepi Menggunakan Delphi 7*. Universitas Gunadarma: Depok.
- Cunningham, P. And Delany, S. J. 2007. "K-Nearest Neighbour Classifier", Issue Technical Report, UCD-CSI-4.
- Kadir, Abdul. 2013. *Dasar Pengolahan Citra dengan Delphi* (Dewiberta Hardjono, Ed.). Yogyakarta: Andi.
- Kadir, A., dan Andhi S. 2013. *Teori dan Aplikasi Pengolahan Citra*. Yogyakarta: Andi.
- Lalujan, T.F. Tanpa tahun. *Mudah Menulis dan Membaca Huruf Hiragana dan Huruf Katakana dalam Bahasa Jepang*. (Online). Tersedia http://www.academia.edu/6228224/menulis_dan_membaca_huruf_hiragana_dan_huruf_katakana_pemula, diunduh 2 April 2017.
- Mukhooyar, Zaky. 2015. *Pengenalan Kata Aksara Jawa Menggunakan Algoritma K-Nearest Neighbor*. Universitas Dian Nuswantoro Semarang: Semarang.
- Munir, Rinaldi. 2004. *Pengolahan Citra Digital dengan Pendekatan Algoritmik*. Informatika: Bandung.
- Nana, F., Budhi, I., S.Si., M.T., Umar, A. A., S.T., M.T. 2016. *Perancangan dan Implementasi Histogram of Oriented Gradients dan K-Nearest Neighbour untuk Deteksi Huruf Hiragana Jepang pada Aplikasi Mobile Penerjemah Kata dalam Bahasa Jepang ke Bahasa Indonesia Berbasis Android*. Institut Teknologi Telkom: Bandung.
- Nugroho, W.E.N. 2016. *Pengenalan Pola Huruf Jepang Hiragana Menggunakan Algoritma Backpropagation*. Skripsi. Yogyakarta: Fakultas Sains dan Teknologi Universitas Sanata Dharma.
- Putra, Darma. 2010. *Pengolahan Citra Digital*. Yogyakarta: Andi.
- Sanjaya, A., M.Kom., Dr.Emma, U., M.Kom., Armadyah, A., M.Kom. 2014. *Identifikasi Personal Berdasar Bentuk Tangan*. Universitas Negeri PGRI Kediri: Kediri.
- Wu, X., Kumar, V. And Quinlan, J.R. 2007. *Top 10 Algorithm in Data Mining*. XIV:1-37.

