

Pengenalan Pola Tulisan Huruf Jepang (Hiragana) Menggunakan Partisi Citra

Ariska Fitria Anggelina¹, Ardi Sanjaya², Ahmad Bagus Setawan³

^{1,2,3}Teknik Informatika, Fakultas Teknik, Universitas Nusantara PGRI Kediri

E-mail: ¹riska1896@gmail.com, ²ardisanjaya@unpkediri.ac.id, ³bagus.este@gmail.com

Abstrak— Beberapa orang memiliki kemampuan dalam mengenali tulisan tangan seseorang melalui pembelajaran. Misalnya pada tulisan tangan huruf Hiragana. 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. 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 grayscalling 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 citra dapat digunakan untuk mengidentifikasi tulisan tangan huruf Hiragana

Kata Kunci - pengenalan pola, partisi citra, tulisan tangan

Abstract – many people have the ability to recognize someone's handwriting through learning. For example on Hiragana letter handwriting. There may be some constraints found during identification, so that one's hand is hard to recognize. The problems of this research are (1) How to design and build a system that can be used as an introduction to the development of handwriting pattern of basic Hiragana (seion)? (2) How to implement the Partition Image method to perform feature extraction on digital image. This research uses Partition method as image feature extraction, as well as K-Nearest Neighbor method as classification and distance calculation using Euclidean Distance. Before the introduction, the image file was preprocessing process first, grayscalling and prewitt edge detection, then performed partition, then identified. From the test results in the first scenario with training data 60 and testing covers all first week writing as much as 90 letters obtained the accuracy of 6%. In the second scenario with the amount of training data 60 and testing covers all the papers of the second week we get 81% accuracy. In the third scenario with the same training data and testing covers all third week writing obtained 88% accuracy. From the test results, it can be concluded that the K-Nearest Neighbor method and Image partition can be used to identify Hiragana letter handwriting

Keywords— pattern recognition, image partition, handwriting

1. PENDAHULUAN

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 [1]. 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 [2]. 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 [2].

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 [3] telah menggunakan algoritma K-Nearest Neighbour untuk mengenali huruf dan angka dengan menggunakan metode K-Nearest Neighbour untuk mengenali pola tulisan tangan angka Arab (Indian) untuk angka 0-9 diperoleh tingkat akurasi hingga 86%. Penelitian lainnya [4] 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 [5] dengan tingkat akurasi sebesar 86,63%. Penelitian lain yang [6] menggunakan partisi citra untuk mengenali pola huruf pada penelitiannya yang berjudul Optical Character Recognition Menggunakan Partisi Citra. Penelitian tersebut mencoba memberikan alternatif baru untuk proses pengenalan karakter pada data citra. Proses yang digunakan yaitu memisahkan paragraf dengan teknik *cropping*. Kemudian memisahkan baris pada paragraf dengan cara membaca area kosong secara vertikal. Dilanjutkan dengan memisahkan karakter pada masing-masing baris dengan cara membaca area kosong secara horisontal. Setelah karakter berhasil dipisah, masing-masing karakter dipartisi menjadi 64 bagian dan kemudian diambil nilai luas piksel tiap-tiap partisinya dan disimpan sebagai data *training*. Untuk tahap pengenalan karakter, menggunakan pencocokan data *testing* terhadap data *training* dan dicari selisih jarak terpendek dengan *euclidean distance*. Font yang digunakan sebagai data *training* dan *testing* yaitu arial, calibri dan times new romans ukuran 8pt, 10pt, 12pt dan 14pt. Didapati hasil bahwa font times new romans memiliki akurasi rendah 23,44% pada ukuran 10pt dan akurasi tertinggi pada font arial ukuran 14pt yaitu sebesar 95,31%

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).

2. METODE PENELITIAN

Berdasarkan uraian diatas, maka peneliti mencoba membangun aplikasi untuk mengenali tulisan tangan Hiragana menggunakan partisi citra dan jarak euclidean untuk menentukan tingkat kemiripan.

2.1. Landasan Teori

2.1.1 Citra Digital

Citra adalah suatu representasi (gambaran), kemiripan, atau imitasi dari suatu objek. Citra terbagi menjadi dua sifat yaitu citra yang bersifat analog dan citra yang bersifat digital. Citra analog adalah citra yang bersifat continue seperti gambar pada televisi. Sedangkan citra digital adalah citra yang dapat diolah komputer [7]. "citra dibagi ke dalam 2 bentuk, yaitu citra tampak dan citra tak tampak" [8]. Citra tampak adalah citra yang dapat dilihat secara langsung oleh alat-alat optik, misalnya foto, gambar dan lukisan. Sedangkan citra tak tampak adalah citra yang dipresentasikan di dalam fungsi matematis (kontinue dan diskrit), misalnya sebuah file citra (.jpg, .jpeg, .bmp, .png) yang disimpan dalam disk komputer berbentuk deret binary.

Citra digital dibentuk dari sejumlah elemen terbatas, yang masing-masing elemen tersebut memiliki nilai dan koordinat tertentu. Citra digital diartikan sebagai suatu citra dapat didefinisikan sebagai fungsi $f(x, y)$ berukuran M baris dan N kolom, dengan x dan y adalah koordinat spasial dan amplitudo f di titik koordinat (x, y) dinamakan intensitas atau tingkat keabuan dari citra pada titik

tertentu. Apabila nilai x, y dan nilai amplitudo f secara keseluruhan berhingga (*finite*) dan bernilai diskrit maka dapat dikatakan bahwa citra tersebut adalah citra digital [9].

2.1.2 Huruf Hiragana

Huruf Hiragana pada awalnya terdiri dari 48 suku kata, namun simbol untuk suku kata “wi” dan “we” telah dihapuskan oleh Departemen Pendidikan Jepang pada tahun 1946, sehingga menjadi 46 suku kata. Perubahan ini merupakan sebagian dari perbaikan suku kata [2][10]. Fungsi huruf *Hiragana* adalah untuk menuliskan kata-kata asli bahasa Jepang yang bukan serapan bahasa asing dan sebagai *Furigana* yaitu kata yang menerangkan bagaimana membaca sebuah huruf *Kanji*. Huruf *Hiragana* dibagi menjadi dua kelompok, yaitu huruf hiragana dasar dan imbuhan.

あ a	い i	う u	え e	お o
か ka	き ki	く ku	け ke	こ ko
さ sa	し shi	す su	せ se	そ so
た ta	ち chi	つ tsu	て te	と to
な na	に ni	ぬ nu	ね ne	の no
は ha	ひ hi	ふ fu	へ he	ほ ho
ま ma	み mi	む mu	め me	も mo
や ya		ゆ yu		よ yo
ら ra	り ri	る ru	れ re	ろ ro
わ wa				を wo
ん n				

Gambar 1. Huruf Hiragana Dasar

が ga	ぎ gi	ぐ gu	げ ge	ご go
ざ za	じ ji	ず zu	ぜ ze	ぞ zo
だ da	ぢ ji	づ dzu	で de	ど do
ば ba	び bi	ぶ bu	べ be	ぼ bo
ぱ pa	ぴ pi	ぷ pu	ぺ pe	ぽ po

Gambar 3. Huruf Hiragana Imbuhan Dakuon

きゃ kya	きゅ kyu	きょ kyo
しゃ sha	しゅ shu	しょ sho
ちゃ cha	ちゅ chu	ちょ cho
にゃ nya	にゅ nyu	にょ nyo
ひゃ hya	ひゅ hyu	ひょ hyo
みゃ mya	みゅ myu	みょ myo
りゃ rya	りゅ ryu	りょ ryo
ぎゃ gya	ぎゅ gyu	ぎょ gyo
じゃ ja	じゅ ju	じょ jo

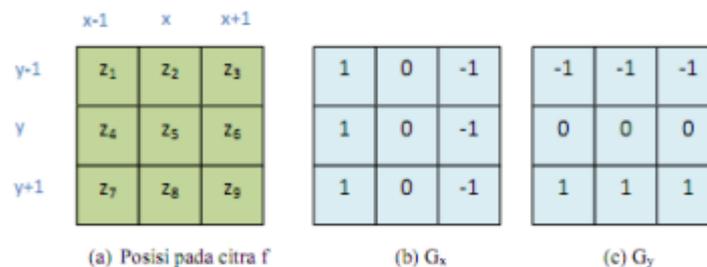
ぢゃ ja	ぢゅ ju	ぢょ jo
びゃ bya	びゅ byu	びょ byo
ぴゃ pya	ぴゅ pyu	ぴょ pyo

Gambar 3. Huruf Hiragana Imbuhan (yoon)

2.1.3 Deteksi Tepi

Edge detection atau deteksi tepi diartikan sebagai berikut : Tepian dari suatu citra mengandung informasi penting dari citra bersangkutan [9]. Tepian citra dapat mempresentasikan objek-objek yang terkandung dalam citra tersebut, bentuk dan ukurannya serta terkadang juga informasi tentang teksturnya. Tepian citra adalah posisi dimana intensitas pixel dari citra berubah dari nilai rendah ke nilai tinggi atau sebaliknya. Deteksi tepi umumnya adalah langkah awal melakukan segmentasi citra. Deteksi tepi berfungsi untuk memperoleh tepi objek. Deteksi tepi memanfaatkan perubahan nilai intensitas yang drastis pada batas dua area. Tepi sesungguhnya mengandung informasi yang sangat penting. Informasi yang diperoleh dapat berupa bentuk maupun ukuran objek. Edge detection dibedakan menjadi dua golongan sebagai berikut : Deteksi tepi dapat dibagi menjadi dua golongan. Golongan pertama disebut deteksi tepi orde pertama, yang bekerja dengan menggunakan turunan atau diferensial orde pertama. Termasuk kelompok ini adalah operator Roberts, Prewitt dan Sobel. Golongan kedua dinamakan deteksi tepi orde kedua, yang menggunakan turunan orde kedua. Contoh yang termasuk kelompok ini adalah Laplacian of Gaussian (LoG) [8].

“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 3x3”[11]. Operator ini mengambil prinsip dari fungsi laplacian yang dikenal sebagai fungsi untuk membangkitkan HPF. Untuk mempercepat komputasi, bagian yang bernilai nol tidak perlu diproses. Bentuk dari operator prewitt ditunjukkan pada gambar berikut :

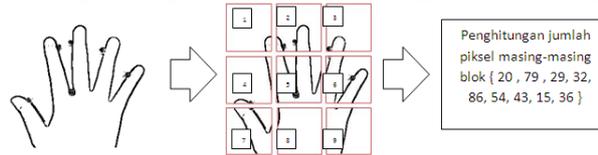


Gambar 4. Operator prewitt (b) dan (c) serta posisi pada citra f

Operator prewitt meletakkan penekanan atau pembobotan pada piksel piksel yang lebih dekat dengan titik pusat dari jendela, dengan demikian ke delapan piksel tetangga mempunyai pengaruh yang sama terhadap perhitungan gradien pada titik pusat jendela.

2.1.4 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 [12].



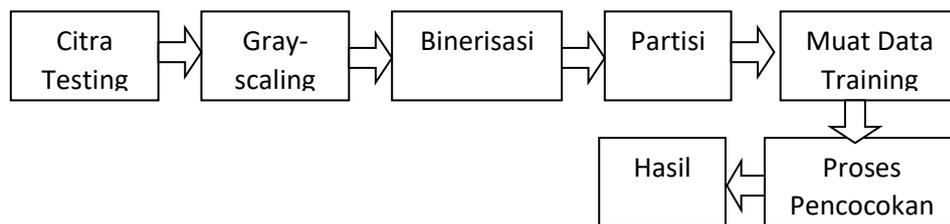
Gambar 5. Contoh partisi citra

2.2 Data Training dan Testing

Data yang digunakan adalah data tulisan tangan huruf hiragana menggunakan kertas putih dan pena berwarna hitam. Diambil dari 90 responden. Proses pengambilan data training dan pengenalan tulisan tangan huruf hiragana disajikan pada gambar 6 dan 7 sebagai berikut :



Gambar 6. Proses Training



Gambar 7. Proses Testing

Saat proses training, tulisan tangan hiragana di scan. Kemudian dilakukan *grayscale* dan binerisasi. Selanjutnya citra di partisi menjadi 9 bagian. Pada masing-masing bagian partisi, diambil nilai binernya dan disimpan sebagai data training. Untuk proses testing juga sama. Saat proses pencocokan, data training di muat dan di cocokkan dengan citra testing untuk masing-masing partisinya. Proses pencocokan menggunakan euclidean distance dan diambil nilai kemiripan terkecil.

2.3 Skenario Pengujian

2.3.1 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.

2.3.2 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.

2.3.3 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.

2.3.4 Skenario uji coba 4

Skenario ini menggunakan data *training* sebanyak 30 citra huruf *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.

3. HASIL DAN PEMBAHASAN

Berikut adalah data hasil skenario pengujian :

- a. 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%.
- b. 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%.
- c. Skenario 3, dari data citra *testing* minggu ketiga sebanyak 90 yang diuji, 74 dikenali dengan benar dan 16 lainnya bernilai salah. Sehingga didapatkan akurasi sebesar 82,22%.
- d. Skenario 4, objek selain huruf *hiragana* ada yang dapat dikenali dan ada yang tidak, sementara objek citra huruf *Hiragana* diluar data huruf yang ada pada data *training* tetap dapat diuji dan dikenali, namun tentu saja pengenalannya bernilai salah.

Tabel 1. Hasil Pengujian

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	6	4	60,00%
Rata-Rata				69,17%

4. SIMPULAN

Dari uji coba yang telah dilakukan pada sistem, dapat ditarik kesimpulan sebagai berikut :

- a. 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.
- b. 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.
- c. Rata-rata akurasi dari semua pengujian adalah 69.17%.

5. SARAN

Untuk penelitian selanjutnya disarankan menambahkan proses penipisan (*thinning*) pada *preprocessingnya*.

DAFTAR PUSTAKA

[1] 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.

- [2] Luluhan, T.F. _____. 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.
- [3] Akbar, R. dan Sarwoko, A.E. 2016. *Studi Analisis Pengenalan Pola Tulisan Tangan Angka Arabik (Indian) Menggunakan Metode K-Nearest Neighbors dan Connected Component Labeling*. Jurnal Dinamika Rekayasa, 12 (2): 45-51.
- [4] Mukhoyyar, Zaky. 2015. *Pengenalan Kata Aksara Jawa Menggunakan Algoritma K-Nearest Neighbor*. Universitas Dian Nuswantoro Semarang: Semarang.
- [5] Nugroho, W.E.N. 2016. *Pengenalan Pola Huruf Jepang Hiragana Menggunakan Algoritma Backpropagation*. Skripsi. Yogyakarta: Fakultas Sains dan Teknologi Universitas Sanata Dharma. Oki. Renshu : Pelajaran Suku Kata Bahasa Jepang (semester 1). Surabaya: Bintang Surabaya
- [6] Sanjaya, A. 2015, *Optical Character Recognition Menggunakan Partisi Citra*. Jurnal Nusantara Of Engineering, Vol. 1/ No. 2/ISSN: 2355-6684
- [7] Sutoyo, T. Mulayanto, E. Vincent S, Oky Dwi N, Wijanarto. 2009, *Teori Pengolahan Citra Digital*, Andi Yogyakarta.
- [8] Kadir, Abdul. 2013. *Dasar Pengolahan Citra dengan Delphi* (Dewiberta Hardjono, Ed.). Yogyakarta:Andi.
- [9] Putra, Darma. 2010. *Pengolahan Citra Digital*. Yogyakarta: Andi.
- [10] Adimihardja, Mulyana. 2011. *Yasashii Nihongo: Bahasa Jepang untk Kelas XI SMA/MA Program Pilihan* (F.M. Erly, Eds.). Bandung: Grafindo Media Pratama.
- [11] Cunningham, P. And Delany, S. J. 2007. "K-Nearest Neighbour Classifier", Issue Technical Report, UCD-CSI-4. Gonzales, R. C., Woods, R. E. 1993. *Digital Images Processing*. Wesley Publishing Company: USA.
- [12] Sanjaya, A., M.Kom., Dr.Emma, U., M.Kom., Armadyah, A., M.Kom. 2014. *Identifikasi Personal Berdasar Bentuk Tangan*. Prosiding Seminar Teknomedia STMIK Amikom Yogyakarta