

Identifikasi Tulisan Tangan Huruf Katakana Jepang Dengan Metode Euclidean

Imam Riadi¹, Abdul Fadlil², Putri Annisa³

¹Program Studi Sistem Informasi, Universitas Ahmad Dahlan

¹Program Studi Teknik Elektro, Universitas Ahmad Dahlan

¹Program Studi Magister Teknik Informatika, Universitas Ahmad Dahlan

Jl. Prof. Soepomo Sh, Umbulharjo, Yogyakarta, Telp. (0274) 563515/Fax. (0274) 31556

imam.riadi@is.uad.ac.id, fadlil@mti.uad.ac.id, putri1807048005@webmail.uad.ac.id

Abstract

Katakana is one of the traditional Japanese letters used to absorption words from other languanges. In the introduction of an object a learning process is needed, which is obtained through the characteristics and experience of observing similar objects after being acquired. But manually it is quite difficult to distinguish between 5 hiragana vowels starting from the image data acquisition process, image processing, feature extraction using Gray Level Co-occurrence Matrix (GLCM) while classifiers use the euclidean distance method. The results of the tests carried out showed an accuracy rate of around 78% using the euclidean method.

Keywords: Katakana, Gray Level Co-occurrence Matrix, Euclidean

Abstrak

Hiragana merupakan salah satu huruf tradisional Jepang yang digunakan untuk menuliskan kata-kata serapan dari bahasa lain. Dalam pengenalan suatu objek dibutuhkan proses pembelajaran, yang didapatkan melalui ciri-ciri dan pengalaman dalam mengamati objek yang serupa setelah didapatkan. Namun secara manual cukup sulit untuk membedakan huruf-huruf tersebut. Tulisan ini menjelaskan sistem identifikasi untuk membedakan antara 5 huruf vocal hiragana yang dimulai dari proses akuisisi data citra, pemrosesan citra, ekstraksi ciri menggunakan Gray Level Co-occurrence Matrix (GLCM) sedangkan pengklasifikasi menggunakan metode jarak Euclidean. Hasil pengujian yang dilakukan menunjukkan tingkat akurasi sekitar 78% menggunakan metode euclidean

Kata kunci: Katakana, Gray Level Co-occurrence Matrix, Euclidean

1. PENDAHULUAN

Pada zaman teknologi sekarang, semakin banyak penerapan panca indera pada komputer. Sistem yang dibuat di komputer dapat mengenali suatu objek dengan mengidentifikasi fungsi, karakteristik dan ciri-ciri dari objek tersebut. Objek sampel sebelumnya telah diidentifikasi menjadi acuan, kemudian melalui pelatihan pengenalan maka objek lainnya diolah oleh algoritma di program komputer. Hal ini dimaksudkan jika ada hal yang tidak dapat diidentifikasi oleh manusia, maka dapat dibantu oleh komputer. Contohnya adalah identifikasi sidik jari, pola tulisan tangan, wajah, dan lain sebagainya. Kebudayaan Jepang saat ini banyak sekali diminati di Indonesia baik itu yang disajikan melalui *manga*, *anime*, film, musik, dan sebagainya. Maka dari itu banyak masyarakat Indonesia yang tertarik mempelajari bahasa Jepang. Bahasa Jepang memiliki hubungan sejarah yang erat dengan negara-negara sekitarnya, salah satunya Cina [1]. Namun karena bahasa ini memiliki aturan gramatical, jenis huruf, cara baca dan cara menulis yang

berbeda-beda, sering kali para pelajar merasa kesulitan dan malas untuk mempelajarinya [2-5]. Terdapat juga jumlah garisan (*stroke*) yang sering ditemui dan menjadikan bahasa Jepang sukar untuk dipelajari. Begitu pula dengan cara penulisan setiap huruf memiliki aturan untuk membentuk sebuah karakter Jepang. Namun manusia memiliki cara penulisan yang berbeda-beda.

Untuk membantu para pelajar dalam mengenali karakter tulisan tangan katakana Jepang, dapat dilakukan dengan menggunakan sistem yang berbasis pengenalan pola. Pengenalan pola merupakan bagian dari kecerdasan buatan, yang mana metode ini memiliki kemampuan untuk mengambil data-data mentah seperti sinyal suara maupun karakter pada citra [6]. Beberapa peneliti telah menggunakan algoritma euclidean untuk mengenali beberapa jenis huruf [7], menggunakan algoritma *Backpropagation Neural Network* untuk melakukan pengenalan karakter, namun juga memakai *Hoar Wavelet* untuk melakukan deteksi lokasi plat nomor dan ekstraksi karakter karakter pada citra plat nomor. Penelitian ini menghasilkan 4 rincian dari citra yaitu rincian dari aproksimasi, vertikal, horizontal, dan diagonal dan diagonal. Keempat rincian ini akan memberikan informasi yang lengkap yang lengkap untuk membantu sistem mendeteksi lokasi plat nomor dan ekstraksi karakter dari citra. Hasil dari metode yang digunakan menunjukkan kelayakan sistem yang diusulkan. Penelitian lainnya juga dilakukan oleh [8], menggunakan metode euclidean untuk pengenalan pola pada citra jamur. Penelitian ini bertujuan untuk membantu petani dalam kegiatan pasca panen yang erat kaitannya dengan mutu produk, dan pada akhirnya menentukan pula harga jual yang dapat diterima oleh petani. Dalam penelitiannya mengungkapkan bahwa nilai rata-rata akurasi yang didapat dengan 30 sampel yang diuji untuk setiap 5 jenis jamur sekitar 78%. Dari masalah tersebut, penulis ingin menguji algoritma euclidean dalam pengenalan pola tulisan tangan yang menjadi kendala dalam pengolahan citra. Penelitian ini bertujuan untuk membantu pelajar dalam mengenali huruf tulisan tangan huruf katakana Jepang dengan bentuk yang tidak biasa dan terkadang sulit untuk dikenali. Citra yang dijadikan acuan adalah 5 huruf vokal untuk menguji beberapa citra lalu melihat presentase keakuratannya.

2. METODOLOGI PENELITIAN

2.1. Alat dan Bahan yang Digunakan

Objek atau bahan yang digunakan untuk diidentifikasi oleh sistem adalah 5 jenis huruf katakana dasar. Masing-masing jenis huruf didapat dari hasil tulisan tangan melalui digital yang berjumlah total 50 data citra, baik untuk citra standar maupun citra uji. Sedangkan alat yang digunakan terbagi atas dua macam yaitu sebagai berikut :

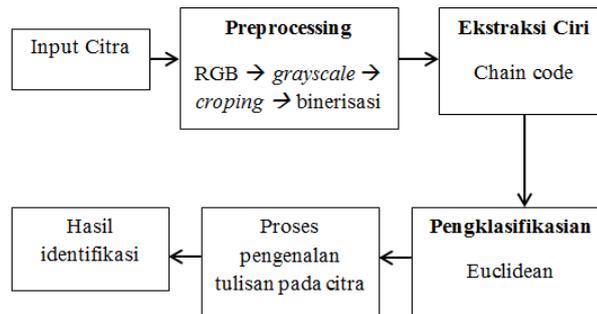
- a) Komputer PC dengan spesifikasi :
 1. Processor : AMD APU A-9 9400 up to 3.2 Ghz
 2. Memory : RAM 4 GB DDR4
 3. Kartu grafis : *Radeon R5 M430 2GB Dedicated VRAM*

b) Sistem yang dibuat untuk mengidentifikasi huruf digunakan perangkat lunak sebagai berikut :

1. Sistem operasi *Microsoft Windows 10*
2. *Software MATLAB R2017a*

2.2. Skema Penerapan Sistem

Adapun gambaran umum yang diusulkan untuk mengenali karkater pada citra yang dilakukan pada penelitian ini meliputi :



Gambar 1. Skema penerapan sistem

Skema dalam sistem ini adalah gambaran untuk membagi langkah-langkah dalam beberapa bagian. Masing-masing bagian harus lengkap tugasnya sebelum berpindah pada proses selanjutnya. Tahapan ini berlaku untuk memproses data citra menjadi data acuan maupun setelah sistem siap pakai untuk diuji oleh pengguna. Diagram di atas terdiri dari beberapa tahap, yaitu :

a) *Input Citra*

Pada tahap ini berfungsi untuk memasukkan citra/gambar tulisan tangan huruf katakana dengan format .jpg untuk diidentifikasi oleh sistem. Citra yang dimasukkan harus sudah tersedia dalam komputer.

b) *Preprocessing*

Pada tahap ini berfungsi untuk melakukan penajaman objek tulisan pada citra agar memudahkan sistem untuk digunakan pada proses-proses berikutnya. Proses awal, citra berwarna RGB (*Red, Green, Blue*) diubah menjadi format citra keabuan (*grayscale*) (7). Proses ini dilakukan dengan rumus berikut :

$$W_{grayscale} = \frac{(R+G+B)}{3} \quad (1)$$

Berikutnya sistem akan melakukan proses binerisasi pada citra yang membuat citra menjadi hanya memiliki 2 warna yaitu hitam dan putih. Pikel-pikel objek menjadi warna hitam yang memiliki nilai 1, dan latar belakang menjadi warna putih yang memiliki nilai 0 (8). Hal ini untuk memudahkan sistem melakukan proses ekstraksi ciri berdasarkan terang

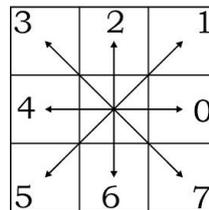
gelapnya suatu citra. Berikut ini contoh hasil citra pada tahap prapengolahan ditunjukkan pada tabel 1.

Tabel 1. Citra pada prapengolahan

Citra Asli	Grayscale	Biner
		
		
		
		
		

c) Ekstraksi Ciri

Pada tahap ekstraksi ciri yang digunakan adalah *chain code*. *Chain code* merupakan metode yang melakukan penelusuran terhadap pixel-pixel objek dengan panduan arah mata angin seperti yang ditunjukkan pada gambar 2.



Gambar 2. Arah mata angin chain code

Mekanisme dari metode ini melakukan penelusuran per-piksel yang digunakan untuk menemukan struktur pembentuk dari suatu objek. Hasil akhirnya berupa sebuah vector ciri yang berisi informasi urutan kode *chain code* pembentuk huruf. Dari proses yang dilakukan maka urutan *chain code* yang dihasilkan untuk setiap huruf dapat memiliki panjang yang berbeda. Tujuan dari proses ini untuk mendapatkan nilai-nilai unik piksel dari citra. Hal ini berfungsi untuk menajamkan pola huruf pada citra dengan *background* sehingga memudahkan sistem untuk memisahkan kategori objek pada proses klasifikasi menggunakan *chain code*.

d) Pengklasifikasian

Proses klasifikasi jarak berfungsi untuk mengelompokkan hasil ekstraksi ciri sehingga diperoleh suatu identifikasi dengan mencocokkan

antar citra. Dengan kata lain, proses ini merupakan proses penajaman data agar mudah digunakan pada proses-proses selanjutnya (9). Metode yang digunakan untuk proses klasifikasi jarak adalah euclidean. Euclidean merupakan suatu metode pencarian kedekatan nilai jarak dari 2 buah variabel serta merepresentasikan tingkat kemiripan antara dua buah citra. Semakin kecil jarak antara dua buah citra maka akan semakin mirip kedua citra tersebut. Pada tahap ini sistem mengelompokkan hasil ekstraksi ciri sehingga diperoleh suatu identifikasi dari citra tersebut menggunakan metode euclidean dengan merepresentasikan citra ke dalam koordinat 2 dimensi (x, y) menggunakan rumus berikut.

$$d_E(x, y) = \sqrt{\sum_{i=1}^d (x_i - y_i)^2} \quad (2)$$

Jarak pada euclidean dihitung berdasarkan ciri khusus yang dimiliki oleh suatu citra, misalnya ciri bentuk. Sebagai contoh terdapat 2 vektor ciri sebagai berikut:

$$A = [0, 3, 4, 5]$$

$$B = [7, 6, 3, -1]$$

Maka, euclidean dari vektor A dan B adalah:

$$\begin{aligned} d_{AB} &= \sqrt{(0 - 7)^2 + (3 - 6)^2 + (4 - 3)^2 + (5 - (-1))^2} \\ &= \sqrt{49 + 9 + 1 + 36} = 9.747 \end{aligned}$$

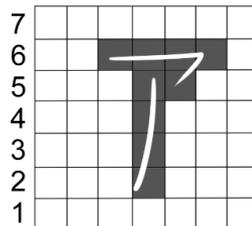
3. HASIL DAN PEMBAHASAN

Pada pengenalan huruf tetap menampilkan citra dan hasil berupa teks bacaan dari citra yang diproses. Hal ini karena pada dasarnya sistem identifikasi dapat diuji setelah sistem identifikasi dapat diuji setelah sistem tersebut dilatih terlebih dahulu [10]. Proses pengujian dilakukan dengan memasukkan citra baru yang belum dikenali atau yang belum pernah dipakai sebelumnya. Hal ini dilakukan untuk mengidentifikasi citra uji dengan citra acuan dan dilihat hasil tingkat kemiripannya. Citra standar dan citra uji disiapkan dengan jumlah yang sama untuk masing-masing jenis huruf katakana. Berikut ini citra yang digunakan pada sistem ditunjukkan pada Tabel 2.

Tabel 2. Jumlah data yang digunakan

Citra Huruf	Jumlah Citra	Standar	Pengujian
A	50	20	30
I	50	20	30
U	50	20	30
E	50	20	30
O	50	20	30

Pada tahap pertama yaitu pemrosesan awal dengan cara memasukkan citra ke dalam proses identifikasi citra huruf katakana dilakukan dengan mengambil citra baru dengan ukuran yang sama dengan citra acuan yang telah tersimpan di komputer. Kemudian sistem akan memproses citra menjadi citra keabuan (*grayscale*) dan citra biner yang bernilai hitam putih. Proses berikutnya masuk pada tahap ekstraksi ciri citra menggunakan *chain code* dengan melakukan pemetaan array sesuai arah mata angin. Berikut ini salah satu huruf A katakana yang sudah dilakukan ekstraksi ciri menggunakan *chain code* ditunjukkan pada gambar 3. berikut.



Gambar 3. Hasil ekstraksi ciri huruf A

Maka plot ke dalam format biner seperti berikut:

```

0   0   0   0   0   0   0
0   0   1   1   1   1   0
0   0   0   1   1   0   0
0   0   0   1   0   0   0
0   0   0   1   0   0   0
0   0   0   1   0   0   0
0   0   0   0   0   0   0
    
```

```

Chain code       : 0 0 0 0 5 4 6 6 6
First difference : 2 0 0 0 5 7 2 0 0
Shape number     : 0 0 0 5 7 2 0 0 2
    
```

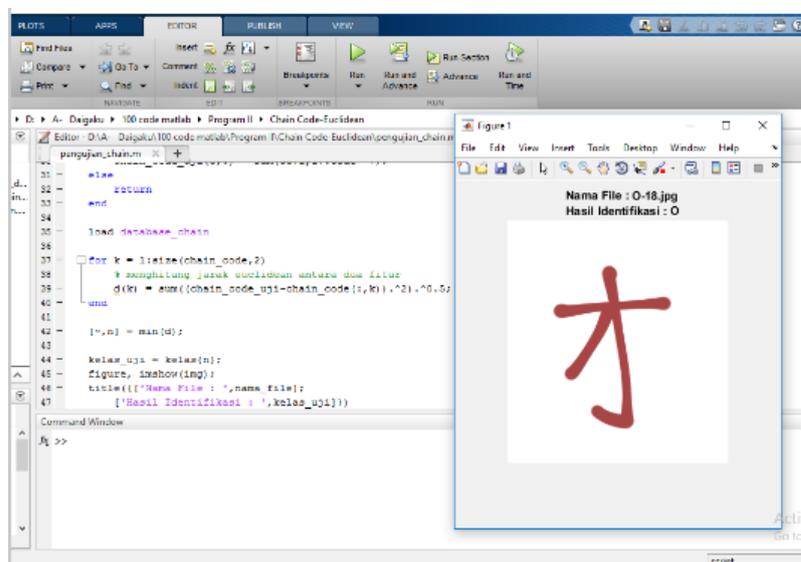
Dari contoh hasil perhitungan penerapan ekstraksi ciri *chain code* diatas, didapat nilai-nilai unik untuk membantu sistem mencocokkan citra uji dengan citra standar melalui penerapan klasifikasi jarak euclidean. Setelah dilakukan pengujian pada setiap citra uji, didapatkan hasil tingkat akurasi seperti yang ditunjukkan pada tabel 3. berikut:

Tabel 3. Hasil pengenalan huruf hiragana

Output/Input	A	I	U	E	O
A	23	4	1	0	2
I	2	26	2	0	0
U	0	1	24	4	1
E	0	2	4	24	0
O	5	2	0	2	21
Akurasi (%) = 120/150*100 = 78,66%					

Diketahui dari tabel 3. bahwa citra uji huruf A dapat mengenali karakter yang ada sebagai huruf A sebanyak 23 citra, dikenali sebagai huruf I sebanyak 4 citra, dikenali sebagai huruf U hanya 1 citra, dan dikenali sebagai huruf O sebanyak 2 citra. Pada huruf I dapat mengenali karakter yang ada sebagai huruf A sebanyak 26 citra, dikenali sebagai huruf A sebanyak 2 citra, dan dikenali sebagai huruf U sebanyak 2 citra. Pada huruf U dapat mengenali karakter yang ada sebagai huruf U sebanyak 24 citra, dikenali sebagai huruf I sebanyak 1 citra, dikenali sebagai huruf E sebanyak 4 citra, dan dikenali sebagai huruf O sebanyak 1 citra. Pada huruf E dapat mengenali karakter yang ada sebagai huruf E sebanyak 24 citra, dikenali sebagai huruf I sebanyak 2 citra, dan dikenali sebagai huruf U hanya 4 citra. Sedangkan pada huruf O dapat mengenali karakter yang ada sebagai huruf O sebanyak 21 citra, dikenali sebagai huruf A sebanyak 5 citra, dikenali sebagai huruf I sebanyak 2 citra, dan dikenali sebagai huruf E sebanyak 2 citra.

Berdasarkan hasil di atas, maka didapatkan rata-rata persentase keberhasilan yaitu sekitar 78%. Tingkat keberhasilan pengenalan yang dihasilkan cukup tinggi meskipun ukuran dan jenis huruf yang digunakan sebagai masukan berbeda dengan citra acuan. Proses ini akan menampilkan identifikasi kemiripan antara dua buah citra yakni citra uji dan citra acuan, kemudian sistem akan menampilkan hasil deteksi citra dikenali sebagai huruf apa oleh sistem. Berikut ini salah satu hasil uji identifikasi citra seperti pada gambar 4.



Gambar 4. Hasil identifikasi huruf O

Pada gambar menunjukkan bahwa sistem dirancang menggunakan Matlab dapat mengenali karakter pada citra dengan baik. Perbandingan dengan penelitian sebelumnya oleh Sapriani Gustani yang mana juga menggunakan metode euclidean untuk proses klasifikasi jaraknya, hasil akurasi yang didapat juga sekitar 78%.

Perbedaan pada penelitian sebelumnya, data yang dipakai merupakan citra foto jamur berjumlah 30 data, sedangkan penelitian kali ini merupakan citra huruf katakana Jepang. Meskipun jenis citra yang digunakan berbeda namun hasil tingkat akurasi tidak jauh berbeda. Hal ini membuktikan bahwa metode euclidean dapat bekerja dengan baik pada citra foto maupun citra berupa tulisan tangan.

Disamping itu, secara keseluruhan hasil penelitian untuk pengenalan pola huruf katakana cukup sesuai dengan penelitian-penelitian sebelumnya karena jumlah data yang berbeda namun dengan selisih persentase tingkat keberhasilan yang mirip.

4. SIMPULAN

Berdasarkan hasil penelitian di atas, maka tujuan penelitian ini yaitu untuk meneliti tingkat keberhasilan yang diterapkan pada huruf katakana Jepang yaitu sekitar 78% bisa dikatakan cukup sesuai. Hal ini membuktikan bahwa metode ekstraksi ciri *chain code* dan klasifikasi jarak euclidean dapat mengidentifikasi huruf, sehingga para pelajar dapat melatih menulis dan mengenal huruf tersebut lebih mudah.

Untuk prospek kajian berikutnya, sistem membutuhkan data citra standar dan data uji lebih banyak. Karena pada dasarnya metode yang dipakai untuk pengolahan citra sangat bergantung pada jumlah citra yang dipakai, oleh sebab itu sistem akan mempunyai *database* yang lebih banyak dan membuat sistem lebih baik dalam mengenal suatu karakter dari citra. Namun citra yang dipakai sebaiknya data primer yang mana citra tersebut bersih dari *noise background* sehingga memudahkan sistem untuk mengolah citra asli menjadi citra biner dan semakin mudah pula dalam proses berikutnya.

DAFTAR PUSTAKA

- [1] Darmawan, Deni, "Inovasi Pendidikan", Remaja Rosdakarya, Jakarta, 2014.
- [2] Sapriani G., A. Fadlil dan R. Umar, "Sistem identifikasi Jamur menggunakan Metode Ekstraksi Ciri Statistik Orde 1 dan Klasifikasi Jarak", Techni.COM, Vol. 16 (4), pp. 378-386, November 2017.
- [3] Sudjianto, "Gramatica Bahasa Jepang Modern", Kesaint Blanc, Jakarta, 2000.
- [4] Parastuti, "Metode Praktis Menguasai Kanji", Genta Group Production, Surabaya, 2015.
- [5] Aditama, Oryza dan Arini Y. A, "Sehari Kuasai Bahasa Jepang", Citra Media, Yogyakarta, 2011.
- [6] Ersya T. dan Youllia I. N, "Implementasi Metode Pattern Recognition untuk Pengenalan Ucapan Huruf Hijaiyyah", JITTER, Vol. 4, No. 1, Desember 2017.
- [7] Saifudin dan A. Fadlil, "Sistem Identifikasi Citra Kayu Berdasarkan Tekstur menggunakan *Gray Level Coocurrence Matrix* (GLCM) dengan Klasifikasi Jarak Euclidean", SINERGI, Vol, 19 No. 3, pp 181-186, Oktober 2015.
- [8] Chairisni L, Tony dan Ardy K., "Pengenalan Karakter Mandarin dengan *Backpropagation Neural Network*", Seminar Nasional Teknologi Informasi dan Multimedia, 4.9-25, Februari 2016.

- [9] Zaitun, "Sistem Identifikasi dan Pengenalan Pola Citra Tanda Tangan menggunakan Sistem Jaringan Saraf Tiruan (*Artificial Neural Network*) dengan Metode *Backpropagation*. Teori dan Aplikasi Fisika, Vol, 3 (2), pp 93-101, July 2015.
- [10] Fitri M, dan A. Fadlil, "Sistem Pengenalan Bunga Berbasis Pengolahan Citra dan Klasifikasi Jarak *Euclidean*", JITEKI, Vol, 3 No. 2, pp 124-131, Desember 2017.