

## Pengenalan Karakter Kanji Menggunakan Fitur Chain Code dan Euclidean Distance

### *Recognition Of Hanzi Characters Using Chain Code And Euclidean Distance Features*

Karman Surya, [suryakarman@gmail.com](mailto:suryakarman@gmail.com)<sup>1)</sup>, I Gusti Ngurah Suryantara, [gusti@bundamulia.ac.id](mailto:gusti@bundamulia.ac.id)<sup>2)</sup>

<sup>1) 2)</sup> Teknik Informatika, Universitas Bunda Mulia

#### ABSTRACT

Mandarin is are the language in China. Mandarin has been widely used, one of which is Indonesia. However, to learn Mandarin Chinese beginners will have difficulty because there are many different characters and have their meanings. Therefore, it is necessary to have a system or application that can recognize or detect Mandarin in order to make it easier for ordinary people to learn Mandarin.

This research is intended to design a system or application that can perform character recognition in Mandarin vocabulary. For systems or applications designed to use a chain code algorithm for pattern recognition in mandarin vocabulary images and for distance calculations using Euclidean distance.

The process that will be carried out in this study is to compare the distance of the mandarin vocabulary image that will be tested with the mandarin vocabulary image that has been trained to get the smallest value of the Euclidean distance.

The results obtained from the tests that have been carried out failed the percentage of success obtained by 34% in recognizing the image of mandarin vocabulary and 66% experiencing failure in recognizing the image of mandarin vocabulary. The failure is caused by the noise contained in the image and does not use the thinning method so that the image of mandarin words that have different thicknesses will produce different feature values as well..

**Keywords:** Chain Code, Feature Extraction, Euclidean Distance, Mandarin, OCR.

#### ABSTRAK

Bahasa Mandarin merupakan bahasa dari negara China. Bahasa Mandarin telah banyak dipakai salah satu nya Indonesia. Namun untuk mempelajari bahasa Mandarin pemula akan mengalami kesulitan karena terdapat banyak karakter berbeda dan memiliki arti nya. Maka dari itu perlu ada nya sistem atau aplikasi yang bisa mengenali atau mendeteksi bahasa Mandarin agar dapat memudahkan orang awam mempelajari bahasa mandarin.

Penelitian ini ditujukan untuk merancang sistem atau aplikasi yang dapat melakukan pengenalan karakter pada kosa kata mandarin. Untuk sistem atau aplikasi yang dirancang memakai algoritma *chain code* untuk pengenalan pola pada citra kosa kata mandarin dan untuk perhitungan jarak menggunakan *euclidean distance*.

Proses yang akan dilakukan di penelitian ini adalah untuk membandingkan jarak citra kosa kata mandarin yang akan dilakukan pengujian dengan citra kosa kata mandarin yang sudah dilatih untuk mendapatkan nilai jarak *euclidean distance* yang terkecil.

Hasil yang didapatkan dari pengujian yang telah dilakukan gagal persentase keberhasilan yang didapatkan sebesar 34% dalam mengenali citra kosa kata mandarin dan 66% mengalami kegagalan dalam mengenali citra kosa kata mandarin. Kegagalan dikarenakan oleh noise yang terdapat pada gambar dan tidak

memakai metode *thinning* sehingga citra kata *mandarin* yang memiliki ketebalan yang berbeda-beda akan menghasilkan nilai fitur yang berbeda-beda juga.

**Kata Kunci:** *Chain Code*, Ekstraksi Fitur, *Euclidean Distance*, *Mandarin*, OCR.

## PENDAHULUAN

### Latar Belakang

Karakter penting untuk dapat menyimpan informasi didalamnya. Pemakaian karakter sendiri bermacam-macam contohnya di negara China, Sistem penulisan huruf China menggunakan *hanzi*. *Hanzi* sendiri memiliki arti pada masing-masing karakternya. Dengan bantuan teknologi kita dapat melakukan pendeteksian karakter dan pengenalan karakter yang akan dapat membantu kita untuk mengetahui arti dari karakter tersebut sehingga dapat memudahkan orang awam untuk mempelajari bahasa mandarin. Pendeteksian karakter dan pengenalan karakter merupakan teknik yang digunakan untuk melakukan proses pengenalan karakter pada komputer.

Pengenalan Karakter Optik [1] adalah proses pengenalan karakter yang menggunakan gambar karakter[2] yang diubah [3] menjadi kode ASCII [4] yang dapat dibaca oleh komputer. Gambar karakter yang digunakan sebagai gambar yang dihasilkan dokumen yang dipindai dan tangkapan layar [5].

Proses pengenalan karakter [6][7] yang akan digunakan adalah memakai *chain code*, pertama citra akan dilakukan deteksi pola dan kemudian akan disimpan ke data base yang nanti akan digunakan untuk perbandingan dengan gambar yang akan diperiksa.

Pada penelitian yang akan dilakukan, penulis akan memakai metode *chain code* untuk melakukan pengenalan pola pada sebuah gambar kosa kata mandarin yang nanti hasil dari metode *chain code* akan dipakai untuk pengenalan karakter.

### Identifikasi Masalah

Berdasarkan permasalahan yang telah dibahas di latar belakang, maka rumusan masalah yang akan dijelaskan pada penelitian ini adalah:

1. Bagaimana cara mengimplementasi dan keakuratan dengan menggunakan algoritma *chain code* pada aplikasi OCR-Mandarin?
2. Bagaimana hasil keakuratan menggunakan algoritma *chain code* pada aplikasi OCR-Mandarin?

#### c. Tujuan dan Manfaat Penelitian

Dari penelitian yang akan dilakukan memiliki tujuan sebagai berikut:

1. Dapat mengimplementasi algoritma *chain code* pada aplikasi OCR Mandarin.
2. Dapat menghasilkan pengenalan citra yang akurat.
3. Membantu orang awam untuk mempelajari bahasa mandarin.

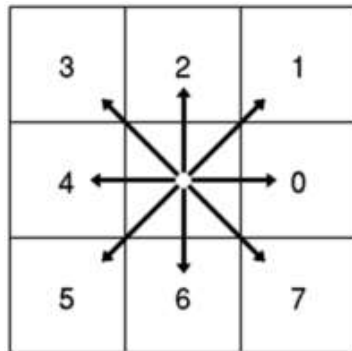
Dari penelitian yang akan dilakukan didapat pula manfaat sebagai berikut:

1. Dapat mengetahui apakah penggunaan algoritma *chain code* bisa mendapatkan hasil yang akurat pada *optical character recognition*.

## METODOLOGI PENELITIAN

### Algoritma *Chain Code*

Kode rantai (*chain code*) dipakai agar dapat mengodekan bentuk atau kontur[6] [8] dari sebuah gambar [9].



**Gambar1.** Arah Kode Rantai

Urutan untuk membaca arah [10] satu titik dari titik yang lain berdasarkan arah jarum jam [11].

### Euclidean Distance

Euclidean Distance mempertimbangkan dua buah vektor fitur [12] yang dibandingkan [13] satu sama lain dengan menghitung jarak [14] antara mereka, atau sebaliknya, menentukan derajat kesamaannya. Ada banyak pengukuran jarak yang digunakan dalam klasifikasi pola visual. Apabila dua vektor fitur  $a = (a_1, a_2, \dots, a_n)$  dan  $b = (b_1, b_2, \dots, b_n)$ , berikut adalah persamaan untuk mengukur jarak yang paling banyak digunakan [15][16].

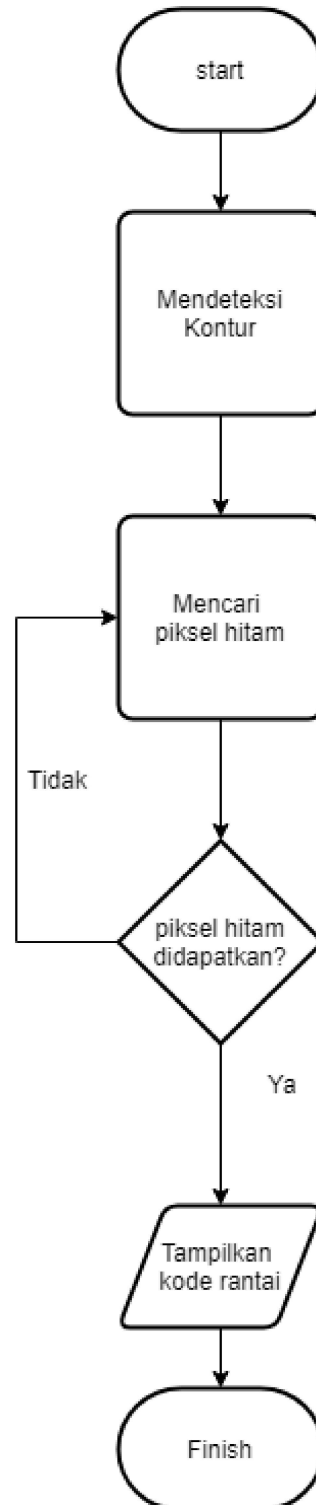
$$d_E = \sqrt{\sum_{i=1}^n (a_i + b_i)^2} \dots\dots\dots (1)$$

dimana  $d_E$  = jarak *eculdean distance*,  $n$  = jumlah vektor,  $a_i$  = vektor citra input,  $b_i$  = vektor citra pembanding.

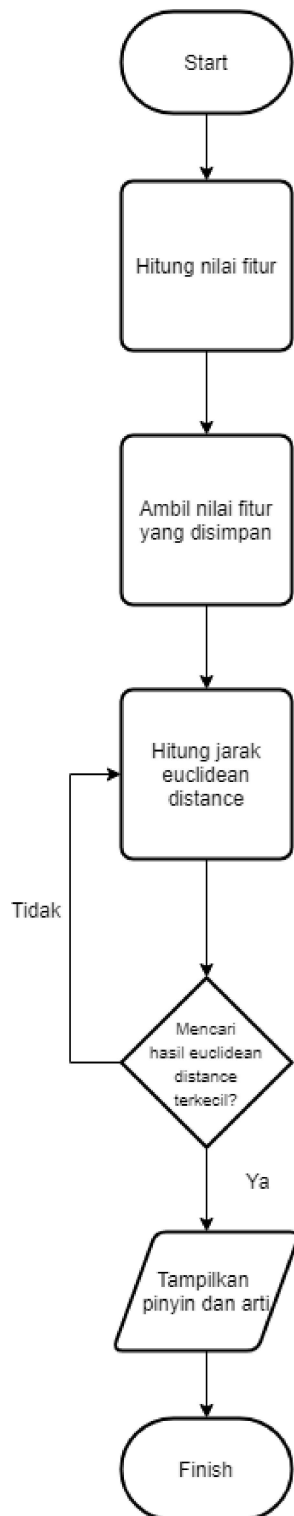
### HASIL DAN PEMBAHASAN

Pada gambar 2 menunjukan proses melakukan pencarian *chain code*. gambar pertama-tama akan dideteksi bentuk outlinenya kemudian akan mencari titik awal atau posisi piksel hitam pertama hingga dapat, kemudian akan mencari

piksel hitam yang terdapat disekitarnya dan ke posisi awal atau pertama.



**Gambar2.** Flowchart Proses Pencarian Chain Code



**Gambar 3** Flowchart Algoritma Euclidean Distance

Pada gambar 3 menunjukan cara kerja algoritma *euclidean distance*. Setelah mendapatkan kode rantai maka akan dilakukan perhitungan menggunakan persamaan untuk mendapatkan nilai fitur. Hasil dari perhitungan fitur akan digunakan untuk mencari nilai jarak terkecil menggunakan *euclidean distance*. *euclidean distance* akan melakukan perhitungan jarak antara nilai fitur yang dimiliki dua buah citra, Dimana hasil dari jarak kedua citra ini akan dibandingkan mana nilai jarak terkecilnya. Kemudian nilai terkecil yang sudah didapatkan akan ditampilkan.

Tabel 1 menampilkan data proses pengujian yang akan dilakukan menggunakan citra kata *mandarin* yang memakai tulisan tangan sebanyak 10 orang relawan yang tiap-tiap tulisan tangan mempunyai 25 kata mandarin.

**Table 1** Citra Mandarin Yang Diuji

No	Mandarin	Pinyin	Arti
1	一	Yī	satu
2	你	nǐ	Kamu
3	好	hǎo	Baik
4	吗	ma	Apakah
5	我	wǒ	Saya
6	也	yě	Juga
7	都	dōu	Semua
8	是	shì	Ya
9	写	Xiě	Menulis
10	听	Tīng	Mendengar
11	说	Shuō	Berbicara
12	爱	Ài	Mencintai
13	切	Qiè	Memotong
14	问	Wèn	Bertanya

15	加	Jiā	Menambah
16	扔	Rēng	Melempar
17	画	Huà	Menggambar
18	炸	Zhà	Menggoreng
19	来	Lái	Datang
20	去	Qù	Pergi
21	回	Huí	Pulang
22	吃	Chī	Makan
23	浴	Yù	Mandi
24	走	Zǒu	Berjalan
25	坐	Zuò	Duduk

Pada tabel 2 menunjukkan hasil keakuratan dari 25 citra kosa kata *Mandarin* dari 10 jenis tulisan tangan berbeda-beda adalah 34%. Dapat disimpulkan untuk hasil pengujian gagal karena hasil keakuratan terlalu rendah.

**Table 2.** Hasil Pengujian Citra Mandarin

No	jenis Tulisan Tangan	Berhasil	Gagal	Keakuratan
1	Tulisan Tangan 1	6	19	24%
2	Tulisan Tangan 2	8	17	32%
3	Tulisan Tangan 3	10	15	40%
4	Tulisan Tangan 4	13	12	52%
5	Tulisan Tangan 5	5	20	20%
6	Tulisan Tangan 6	8	17	32%
7	Tulisan Tangan 7	6	19	24%
8	Tulisan Tangan 8	11	14	44%
9	Tulisan Tangan 9	11	14	44%
10	Tulisan Tangan 10	8	17	32%
Rata-rata				34%

## SIMPULAN

Berdasarkan dari penelitian dan pengujian yang telah dilakukan dapat disimpulkan sebagai berikut:

1. *Chain code* dapat diimplemtasikan untuk pengenalan citra kata *mandarin* namun untuk pengujian keakuratan gagal karena hanya mendapatkan persentase keakuratan sebesar 34%. Penyebab kegagalan disebabkan adanya noise pada gambar dan tidak memakai metode *thinning* sehingga citra memiliki ketebalan yang berbeda-beda.
2. Objek dari citra yang memiliki ketebalan dan ukuran yang berbeda dapat mempengaruhi kode rantai dan nilai fitur yang ada dicitra.

## DAFTAR PUSTAKA

- [1] D. Normalasari and I. Afrianto, "Aplikasi Identifikasi Kata Berbasis Optical Character Recognition dan Augmented Reality," *Komputa J. Ilm. Komput. dan Inform.*, vol. 8, no. 2, pp. 77–85, 2019, doi: 10.34010/komputa.v8i2.3053.
- [2] T. M. S. Mulyana and Herlina, "Evenly brightening using kurtosis Gaussian pattern to simplify image binarization," *J. Phys. Conf. Ser.*, vol. 1397, no. 1, p. 012076, Dec. 2019, doi: 10.1088/1742-6596/1397/1/012076.
- [3] R. Agustian, N. A. H., and J. Karel, "Implementasi Metode Chain Code Untuk Pengenalan Rambu Lalu Lintas," *J. Inform.*, vol. 12, no. 1, pp. 67–75, 2016, doi: 10.21460/inf.2016.121.413.
- [4] M. A. Pradinata, A. Prafanto, and E. Budiman, "Vehicle Detection Using Optical Character Recognition," vol. 4, no. 1, pp. 30–34, 2019.
- [5] N. Venkata Rao, A. S. C. S. Sastry, A. S. N. Chakravarthy, and P.

- 
- Kalyanchakravarthi, "Optical character recognition technique algorithms," *J. Theor. Appl. Inf. Technol.*, vol. 83, no. 2, pp. 275–282, 2016.
- [6] T. M. S. Mulyana, D. Widyaningrum, and H. Herlina, "OCR HURUF JAWA DENGAN FITUR KODE RANTAI DAN LEVENSCHTEIN DISTANCE," *Netw. Eng. Res. Oper.*, vol. 6, no. 1, p. 67, Apr. 2021, doi: 10.21107/nero.v6i1.217.
- [7] L. U. De Bourgogne, "Pendekatan kode rantai sebagai dasar pengenalan karakter," vol. 2009, no. Snati, 2009.
- [8] A. N. Fadillah, A. Rahmi, K. Purwandari, I. Srimuddawamah, R. Etika, and E. Candrarini, "Island Shape Detector Menggunakan Chain Code," pp. 3–6.
- [9] E. P. P. Ezy Claudia Nivsky, Ernawati, "Aplikasi Biometrika Pencocokan Citra Daun Telinga Berbasis Tekstur Dan Bentuk Menggunakan Metode Transformasi Wavelet Dan Chain Code. ISSN : 2303-0755," *J. Rekursif*, vol. 4, no. 3, pp. 325–333, 2016.
- [10] S. Rahman and M. Ulfayani, "Perancangan Aplikasi Identifikasi Biometrika Telapak Tangan Menggunakan Metode Freeman Chain Code," *CESS (Journal Comput. Eng. Syst. Sci.)*, vol. 2, no. 2, pp. 64–73, 2017, doi: <https://doi.org/10.24114/cess.v2i2.6195>.
- [11] K. R. R. Wardani, A. Setiawan, and Y. Setiawan, "Penerapan Sistem Pakar Berbasis Aturan dan Metode Freeman Chain Code pada Aplikasi Pengukuran Isi Botol," *J. Telemat.*, vol. 13, no. 2, pp. 67–72, 2018.
- [12] D. Nugraheny, "Metode Nilai Jarak Guna Kesamaan Atau Kemiripan Ciri Suatu Citra (Kasus Deteksi Awan Cumulonimbus Menggunakan Principal Component Analysis)," *Angkasa J. Ilm. Bid. Teknol.*, vol. 7, no. 2, p. 21, 2017, doi: 10.28989/angkasa.v7i2.145.
- [13] M. Nishom, "Perbandingan Akurasi Euclidean Distance, Minkowski Distance, dan Manhattan Distance pada Algoritma K-Means Clustering berbasis Chi-Square," *J. Inform. J. Pengemb. IT*, vol. 4, no. 1, pp. 20–24, 2019, doi: 10.30591/jpit.v4i1.1253.
- [14] N. Wakhidah, "Clustering Menggunakan K-Means Algorithm," *J. Transform.*, vol. 8, no. 1, p. 33, 2010, doi: 10.26623/transformatika.v8i1.45.
- [15] D. Wahiddin, "Klasifikasi Kadar Hidrasi Tubuh Berdasarkan Warna Urine dengan Metode Ekstraksi Fitur Citra dan Euclidean Distance," *Techno Xplore J. Ilmu Komput. dan Teknol. Inf.*, vol. 5, no. 1, pp. 16–20, 2020, doi: 10.36805/technoxplore.v5i1.887.
- [16] T. M. S. Mulyana and A. Harjoko, "A chinese character recognition method based on population matrix and relational database," *Proc. 2nd Int. Conf. Inf. Commun. Technol. Semin. ICTS 2006*, no. 1, pp. 518–523, 2006.