

KLASIFIKASI HURUF KATAKANA DENGAN METODE TEMPLATE MATCHING CORRELATION

Rezki Trianto¹, Ni Nengah Dewi Merdekawati², Rizkie Purnama Shakti Nugraha³, Dita Nur Yuni Astiti⁴, Happy Gagas Tri Atmojo⁵

Jurusan Informatika/Ilmu Komputer, Program Teknologi Informasi dan Ilmu Komputer,
Universitas Brawijaya, Malang, 65145, Indonesia.

E-Mail: rezkitrianto@gmail.com¹, ewiiq.dewiiq@gmail.com², purnamarizkie@gmail.com³, ditashata@gmail.com⁴, xtraper77@gmail.com⁵

Abstrak

Algoritma *Template Matching* merupakan metode sederhana yang digunakan untuk mengenali pola pada sebuah citra. Metode *Template Matching* sering digunakan untuk mengidentifikasi citra karakter huruf, angka, sidik jari dan aplikasi-aplikasi pencocokan citra lainnya. Pada Project Akhir ini dibuat sebuah aplikasi identifikasi huruf jepang khususnya katakana yang digunakan untuk tujuan memperkenalkan bahasa jepang kepada masyarakat luas. Metode yang digunakan pada aplikasi ini adalah *Template Matcing Correlation*, yaitu teknik statistik yang digunakan untuk mencari dua variabel/matriks atau lebih yang sifatnya kuantitatif. Algoritma ini mencocokkan setiap pixel pada suatu matriks citra digital dengan citra yang menjadi template acuan. Hasil dari uji coba menunjukkan bahwa pengenalan pola huruf katakana dengan metode *Template Matching Correlation* ini dapat mendeteksi gambar huruf katakana yang diujikan. Tingkat keberhasilan pengenalan huruf tersebut sebesar 85% dari data uji sebanyak 20 citra.

Kata Kunci: Template Matcing Correlation, Penerjemah, Katakana, Optical Character Recognition.

Abstract

Template Matching algorithm is a simple method that is used to recognize patterns. Template Matching method often used to identify the image of the character letters, numbers, fingerprint and applications matching image. In this Final Project an application specifically Japanese katakana letters identification is made for the purpose of introducing Japanese language to the general public. The method used in this application is Matcing Correlation Template, the statistical technique that is used to quantitatively find the two variables / matrices or more. This algorithm to match each pixel in a digital image with an image matrix using the reference template as matches ground truth. The results of the tests show that the pattern recognition of katakana letters using template Matching Correlation method can detect the image of katakana letters that has been tested. The success rate of identifying this tested katakana letters is at 85% from the data test as much as 20 images.

Key words: Template Matching Correlation, Translator, Katakana, Optical Character Recognition.

PENDAHULUAN

Pengenalan karakter huruf jepang sebagai pembelajaran terhadap bahasa jepang khususnya pada huruf katakana semakin populer dikalangan masyarakat saat ini. Penggunaannya yang sederhana dan tingkat pengenalan huruf yang tinggi akan dapat meningkatkan minat pengguna untuk mempelajari bahasa jepang tersendiri.

Aplikasi pengenalan karakter katakana jepang ini termasuk kedalam jenis *Optical Character Recognition* (OCR) yang merupakan jenis aplikasi untuk mengidentifikasi citra huruf kemudian diubah ke dalam bentuk tulisan.

Sistem pengenalan pola karakter huruf katakana ini menggunakan metode klasifikasi. Pengenalan pola sendiri merupakan suatu ilmu yang digunakan untuk mengklasifikasikan atau menggambarkan sesuatu berdasarkan pengukuran kuantitatif fitur (ciri) atau sifat utama dari sebuah objek. Pola sendiri merupakan komposit/gabungan dari fitur yang merupakan sifat dari sebuah objek [2].

Untuk mengenali pola dari huruf dapat dilakukan proses ekstraksi ciri terlebih dahulu. proses ini digunakan untuk mendapatkan ciri atau fitur khusus dari sebuah data. Pada sebuah citra, fitur-fitur ini dapat berupa pixel pada sebuah matriks yang terbentuk menjadi sebuah citra digital. Proses ekstraksi ciri ini diimplementasi pada proses *pre-processing* yang dilakukan pada sebuah citra. Hal ini penting untuk dilakukan, untuk peningkatan persentasi keberhasilan pencocokan dari sebuah objek. diantaranya seperti dilakukan pengubahan ukuran citra agar ukuran pixel citra yang dibandingkan serupa hingga dilakukan proses *thresholding* yang digunakan untuk menyeragamkan nilai pixel dari citra serta penghilangan *noise* yang ada. Setelah proses ekstraksi ciri telah dilakukan selanjutnya dilakukan proses pengenalan huruf katakana dengan metode pengenalan pola.

Banyak sekali metode pengenalan pola yang dapat digunakan untuk mengenali sebuah pola dari suatu objek, diantaranya seperti *Naive Bayes*, *Linear Discriminant Analysis* (LDA), Jaringan Syaraf Tiruan, *K-mean Clustering*, *Template Matching* dan lain sebagainya. Pada pengenalan karakter katakana ini, digunakan metode klasifikasi dengan memanfaatkan metode *Template*

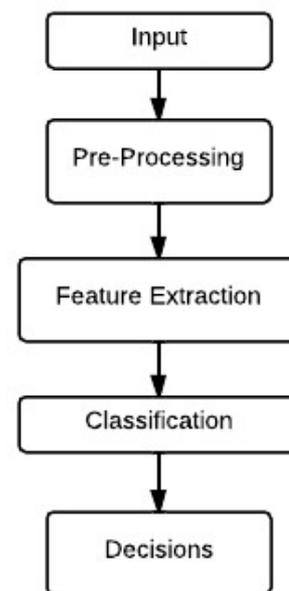
Matching yang merupakan proses membandingkan gambar dengan menggunakan *stored template* [3], salah satunya adalah *Template Matching Correlation*.

Metode klasifikasi *Template Matching Correlation* merupakan algoritma yang sederhana yang mencocokkan masing-masing pixel dari sebuah matriks dengan matriks lainnya. Semakin besar simpangan dari matriks pembandingnya, maka tingkat kemiripan semakin rendah.

Algoritma *Template Matching Correlation* ini merupakan algoritma yang mudah diimplementasikan dan mempunyai persentasi keberhasilan pencocokan karakter yang sangat tinggi.

METODOLOGI

Metode yang digunakan pada pengenalan karakter huruf katakana ini adalah metode *Template Matching Correlation*. Hal-hal yang dilakukan pada pengenalan pola huruf katakana dapat dilihat pada Gambar 1.



Gambar 1. Tahap Pengenalan Huruf

1. Input

Proses *input* pertama kali dilakukan sebagai proses masukan citra yang nantinya akan dilakukan pengenalan huruf. File masukan dapat berupa format *.jpg ataupun *.png,

2. Pre Processing

Proses *Pre-Processing* pada citra digunakan untuk penyamaan ukuran matriks yang akan dicocokkan dengan algoritma *Template Matching Correlation* serta penghilangan *noise* pada citra.

2.1 Resize Image

Pertama kali dilakukan perubahan ukuran gambar menjadi berukuran 15x15 yang dimaksudkan agar ukuran matriks citra yang dicocokkan dengan menggunakan metode *Template Matching Correlation* sama.

2.2. Grayscale

Dilakukan proses *grayscale* untuk mengubah warna citra menjadi keabu-abuan. *Grayscale* merupakan citra digital yang hanya memiliki satu nilai kanal pada setiap pixelnya, dengan kata lain bagian *Red = Green = Blue*. nilai tersebut digunakan untuk menunjukkan tingkat intensitas. Citra *grayscale* dibentuk dari 8 bit per informasi pixel dan menggunakan 256 bayangan warna abu-abu sebagai bagian dasar dari warna [4][5].

Pada proses *grayscale*, masing-masing pixel RGB dari citra diambil nilainya, lalu dilakukan pengambilan mean dari ketiga nilai RGB tersebut, yang nantinya masing-masing nilai R, G, dan B akan diinisialisasi dengan nilai rata - ratanya tersebut sehingga tercipta warna keabu-abuan dari matriks citra yang telah dilakukan proses *grayscale*.

2.3. Thresholding

Thresholding adalah suatu proses yang digunakan untuk menghasilkan citra biner yaitu citra dengan hanya dua warna, yaitu: hitam dan putih. Nilai pixel yang awalnya lebih kecil daripada 127 diinisialisasi dengan nilai pixel 1, sedangkan nilai pixel yang awalnya lebih besar daripada 127 yang berupa *background* diinisialisasi dengan nilai pixel 0.

$$f(\text{threshold}) = \begin{cases} 1, & \text{jika pixel} \leq 127 \\ 0, & \text{jika pixel} > 127 \end{cases} \quad (1)$$

Keterangan :

$f(\text{threshold})$: Nilai pixel hasil thresholding.

pixel : Nilai pixel yang telah dilakukan proses grayscaleing.

Selain digunakan untuk perubahan warna citra menjadi biner, *thresholding* juga digunakan untuk penghilangan *noise* yang dapat mengganggu hasil dari pengenalan karakter nantinya.

3. Feature Extraction

Proses *Feature Extraction* merupakan proses yang dilakukan untuk mendapatkan fitur atau ciri dari gambar yang nantinya akan dilakukan proses klasifikasi dengan menggunakan algoritma *Template Matching Correlation*. Kelas yang ada pada data latih huruf katakana sejumlah 46, dan dapat dilihat pada Gambar 2.

A	I	MI	NU	SE	U
CHI	KA	MO	O	SHI	WA
E	KE	MU	RA	SO	WO
HA	KI	N	RE	SU	YA
HE	KO	NA	RI	TA	YO
HI	KU	NE	RO	TE	YU
HO	MA	NI	RU	TO	
FU	ME	NO	SA	TSU	

Gambar 2. Daftar Kelas

Sedangkan fitur yang digunakan merupakan nilai masing-masing pixel pada citra yang digunakan sebagai data latihnya, yaitu dimulai dari pixel (0,0), (0,1), ... , (14,14) yang total seluruhnya sejumlah 255 fitur.

Untuk mendapatkan data latih, digunakan 5 jenis huruf katakana yang memiliki perbedaan dari jenis hurufnya. Kelima jenis huruf katakana tersebut akan dilakukan proses operasi "*OR*" terhadap masing-masing pixelnya sehingga nantinya akan ditemukan nilai matriks citra baru yang akan dibandingkan dengan data uji yang ada. Contoh data latih dapat dilihat pada Gambar 3.



Gambar 3. Contoh data latih huruf katakana A

4. Classification

Proses *classification* menggunakan metode *Template Matcing Correlation*, algoritma yang sederhana untuk diterapkan pada pengenalan pola citra digital dan memiliki tingkat keakuratan yang tinggi.

Template Matcing sendiri merupakan proses mencari suatu objek pada keseluruhan objek yang berada dalam suatu citra.

Template Matcing Correlation adalah adalah teknik statistik yang digunakan untuk mencari dua variabel/matriks atau lebih yang sifatnya kuantitatif. Algoritma ini biasa digunakan untuk mencocokkan dua matriks citra, dengan akurasi yang cukup besar/tepat. Tetapi akurasi yang besar tersebut terbayar dengan kekurangannya, yaitu penghitungan dan penyimpanan data yang cukup besar untuk mencapai hasil optimum. Algoritma ini mencocokkan setiap pixel pada suatu matriks citra digital dengan citra yang menjadi template (acuan) [1].

Algoritma *Template Matcing Correlation* tentu memiliki kelebihan serta kekurangan. Kelebihannya adalah algoritma ini sangat sederhana untuk diimplementasi kedalam suatu aplikasi berbasis pengolahan citra digital. Algoritma ini juga memiliki tingkat keberhasilan yang sangat tinggi terhadap kemiripannya. Sedangkan untuk kelemahannya, algoritma *Template Matcing Correlation* memiliki perhitungan dan penyimpanan data yang besar untuk mencocokkan dan mendapatkan hasil yang diinginkan. Serta penerapan aplikasi *Optical Character Recognition* membutuhkan data latih yang cukup banyak untuk mendapatkan hasil yang maksimal, dikarenakan semakin banyak jenis huruf yang dilatih maka kemungkinan kemiripan dari huruf yang akan dicocokkan tersebut akan semakin besar.

Algoritma *Template Matcing Correlation* tersebut akan dijabarkan sebagai berikut :

$$r = \frac{\sum_{k=1}^n (x_{ik} - \bar{x}_i) \cdot (x_{jk} - \bar{x}_j)}{\sqrt{\left[\sum_{k=1}^n (x_{ik} - \bar{x}_i)^2 \cdot \sum_{k=1}^n (x_{jk} - \bar{x}_j)^2 \right]}} \quad (2)$$

Dimana, x_i dan x_j merupakan rata-rata dari matriks i dan j yang dapat dihitung dengan :

$$\begin{aligned} \bar{x}_i &= \frac{1}{n} \sum_{k=1}^n x_{ik} \\ \bar{x}_j &= \frac{1}{n} \sum_{k=1}^n x_{jk} \end{aligned} \quad (3)$$

Keterangan :

r = nilai korelasi antara dua buah matriks

x_{ik} = nilai pixel ke- k pada matriks i

x_{jk} = nilai pixel ke- k pada matriks j

\bar{x}_i = rata-rata nilai pixel matriks i

\bar{x}_j = rata-rata nilai pixel matriks j

n = jumlah pixel pada suatu matriks

5. Decisions

Proses Penentuan klasifikasi citra yang telah diuji dengan menggunakan algoritma *Template Matcing Correlation*.

HASIL DAN PEMBAHASAN

Implementasi dari penggunaan algoritma *Template Matcing Correlation* ini digunakan untuk menentukan klasifikasi dari huruf katakana yang telah diinputkan sebelumnya. Input huruf katakana yang telah dilakukan proses ekstraksi fitur akan dihitung nilai korelasinya dengan masing-masing data latih yang telah dilakukan operasi "OR" sebelumnya.

Data uji yang akan digunakan adalah data huruf katakana berupa citra RGB sebanyak 20 contoh citra yang diambil secara acak yang merupakan huruf yang berbeda dari data latihnya dalam segi jenis ataupun ukuran. Pada masing-masing citra yang akan dilakukan uji coba, akan didapatkan nilai korelasinya sesuai dengan persamaan (2) dengan masing-masing data latih yang telah didapat sebelumnya. Contoh data uji dapat dilihat pada Gambar 4 dan *interface* aplikasi ini dapat dilihat pada Gambar 5.

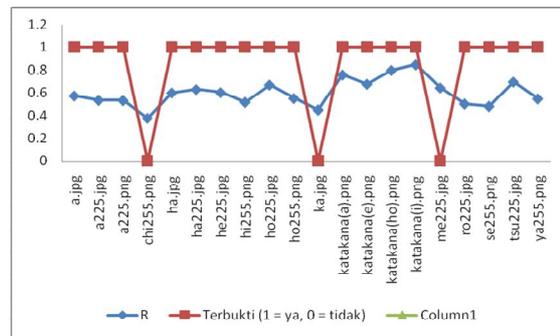


Gambar 4. Data Uji

Tabel 1. Nilai korelasi data uji

Data	Korelasi	
	r	Terbukti
a.jpg	0.5738127	Ya
a225.jpg	0.5355558	Ya
a225.png	0.5339113	Ya
chi255.png	0.3809302	Tidak
ha.jpg	0.6008679	Ya
ha225.jpg	0.630705	Ya
he225.jpg	0.6038715	Ya
hi255.png	0.5183644	Ya
ho225.jpg	0.6712171	Ya
ho255.png	0.5517127	Ya
ka.jpg	0.4531754	Tidak
katakana(a).png	0.7605818	Ya
katakana(e).png	0.6786739	Ya
katakana(ho).png	0.8005685	Ya
katakana(i).png	0.850318	Ya
me225.jpg	0.6448994	Tidak
ro225.jpg	0.5041841	Ya
se255.png	0.4835884	Ya
tsu225.jpg	0.6985561	Ya
ya255.png	0.5481974	Ya

Dari uji coba data latih sebanyak 20 citra, didapat bahwa 85% dari data latih yang diujikan benar. hal ini menunjukkan bahwa pengenalan pola dengan metode *Template Matching Correlation* mempunyai tingkat keberhasilan yang cukup tinggi, serta nilai korelasi yang cukup tinggi yang berkisar antara 0.5 hingga 1 yang membuktikan bahwa dengan nilai korelasi tersebut menunjukkan bahwa korelasi antara matriks pada citra uji dengan matriks pada data latih mempunyai kemiripan yang besar, sedangkan pada nilai korelasi diantara 0 hingga 0.5 memiliki hubungan korelasi tetapi tidak terlalu kuat. Contoh nilai korelasi dapat dilihat pada Tabel 1 dan grafik hubungan dengan hasil dapat dilihat pada Gambar 6.

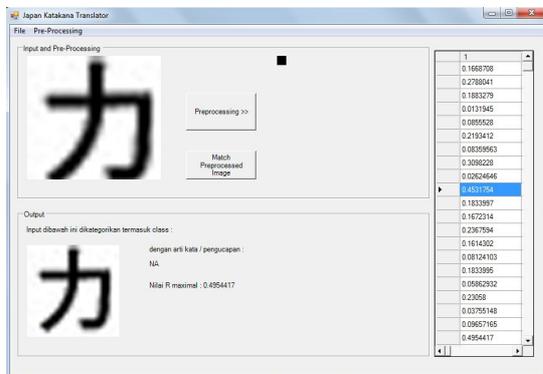


Gambar 6. Grafik hubungan nilai R dengan hasil klasifikasi

Nilai korelasi yang didapatkan dari percobaan nilainya beragam, yang berkisar antara -1 hingga 1, nilai r dapat dinyatakan sebagai berikut :

$$-1 \leq r \leq 1 \quad (4)$$

Kedua matriks dikatakan memiliki korelasi jika nilai korelasi berkisar antara 0 dan 1, jika nilai korelasi sama dengan 0 (nol) maka tidak terdapat hubungan antara kedua buah matriks tersebut, jika nilai korelasinya 1 (satu) maka terdapat hubungan positif yang sempurna, sedangkan jika nilai korelasinya -1 (negatif satu) maka terdapat hubungan linear negatif yang sempurna dari kedua matriks yang dicocokkan dengan menggunakan algoritma korelasi tersebut dan dikatakan berkorelasi negatif. Hubungan positif tersebut terjadi bila kenaikan satu variabel diikuti dengan kenaikan variabel lainnya sedangkan



Gambar 5. Japan Katakana Translator

hubungan negatif diikuti kenaikan satu variabel diikuti dengan penurunan variabel lainnya [6][7]. Semakin besar nilai korelasinya, maka kemiripan matriks yang dibandingkan semakin besar.

KESIMPULAN

Algoritma *Template Matching Correlation* merupakan algoritma yang sederhana untuk diimplementasi dalam mengenali sebuah pola dari citra dengan tingkat keberhasilan yang sangat tinggi dalam mengenali objeknya tetapi membutuhkan memory dan penyimpanan yang cukup besar dalam pemrosesannya.

Algoritma ini mencocokkan masing-masing pixel input dengan data latihnya, jika terdapat perbedaan dalam segi posisi, ataupun jenis huruf, maka akan mempengaruhi hasil dari nilai korelasi tersebut.

Nilai korelasi berkisar antara -1 hingga 1, semakin mendekati nilai 1 maka tingkat kemiripan semakin besar.

DAFTAR PUSTAKA

- [1] Putra, Darma, "Pengolahan Citra Digital", Penerbit Andi, Yogyakarta, 2010
- [2] Fatta, Hanif Al, "Rekayasa Sistem Pengenalan Wajah", Penerbit Andi, Yogyakarta, 2009
- [3] Wilkinson, Barry and Michael Allen, "Parallel Programming - Teknik dan Aplikasi Menggunakan Jaringan Workstation & Komputer Paralel", Penerbit Andi, Yogyakarta, 2010
- [4] Komputer, Wahana, "Teknik Pengolahan Image Bitmap dengan GIMP", Penerbit Andi, Yogyakarta, 2004
- [5] Enterprise, Jubilee, "Mengupas Rahasia Warna Photoshop", Penerbit Elex Media Komputindo, Jakarta, Tanpa Tahun
- [6] Saefuddin, Asep, dkk, "Statistika Dasar", Penerbit Grasindo, Jakarta, Tanpa Tahun
- [7] Supranto, J, "Statistik: Teori & Aplikasi, edisi 6, jilid 1", Penerbit Erlangga, Jakarta, 2000

PENULIS



Rezki Trianto merupakan mahasiswa jurusan Informatika / Ilmu Komputer Universitas Brawijaya, yang memiliki minat terhadap Rekayasa Perangkat Lunak.



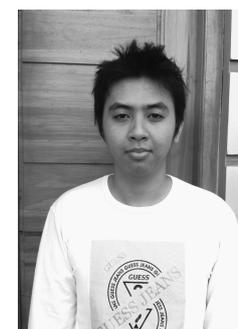
Ni Nengah Dewi Merdekawati merupakan mahasiswa Jurusan Informatika / Ilmu Komputer Universitas Brawijaya, yang memiliki minat terhadap Rekayasa Perangkat Lunak.



Rizkie Purnama merupakan mahasiswa jurusan Informatika / Ilmu Komputer Universitas Brawijaya, yang memiliki minat terhadap Rekayasa Perangkat Lunak.



Dita Nur Yuni Astiti merupakan mahasiswa jurusan Informatika / Ilmu Komputer Universitas Brawijaya, yang memiliki minat terhadap Rekayasa Perangkat Lunak.



Happy Gagas Tri Atmojo merupakan mahasiswa jurusan Teknik Informatika Universitas Brawijaya, yang memiliki minat terhadap Komputasi Cerdas.