

ANALISIS DAN IMPLEMENTASI PENGENALAN HURUF ARAB MENGGUNAKAN MODIFIED DIRECTION FEATURE EXTRACTION DAN LEARNING VECTOR QUANTIZATION

Muhammad Amrul Ummami¹, Adiwijawa², Retno Novi Dayawati³

¹Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

Abstrak

Sistem pengenalan pola saat ini sudah berkembang dan banyak dimanfaatkan . Salah satu dari pengenalan pola yang umum dikenal orang adalah pengenalan tulisan. Huruf Arab merupakan huruf yang wajib dipelajari oleh umat muslim, serta memiliki kemiripan pola pada beberapa karakter sehingga menyebabkan pengenalan pola terhadap huruf Arab memiliki tingkat kesulitan yang tinggi.

Dengan permasalahan tersebut dibangun sistem pengenalan pola yang memiliki kemampuan untuk yang mengenali huruf Arab yang ditulis tangan dengan output berupa pembacaan huruf yang benar dengan menggunakan kombinasi metode Modified Direction Feature (MDF) dan Learning Vector Quantization (LVQ), menganalisis tingkat akurasi dari pengenalan huruf Arab, menganalisis faktor-faktor yang berpengaruh terhadap tingkat akurasi dari pengenalan huruf Arab menggunakan MDF dan LVQ.

Teknik MDF merupakan teknik yang mengambil vektor ciri dari segala arah dan menggabungkannya sehingga terbentuk vektor ciri yang spesifik. Dari vektor ciri yang diambil dari segala arah tersebut dapat diperlihatkan keunikan yang dimiliki oleh huruf Arab untuk inputan LVQ yang mengklasifikasikan pola yang masing-masing unit keluaran mewakili kategori atau kelas tertentu sehingga hasil yang didapatkan menjadi maksimal.

Sistem pengenalan pola huruf Arab dengan menggunakan MDF dan LVQ menghasilkan kesimpulan bahwa teknik MDF dan LVQ dapat digunakan dalam mengekstraksikan pola dan mengklasifikasikan huruf Arab, dimana vektor ciri pola yang dihasilkan dari MDF memiliki nilai yang tidak terlalu tinggi dan dapat memberikan keunikan pada pola, serta pengklasifikasian dari LVQ memiliki ketepatan yang tidak terlalu tinggi. Hal ini disebabkan karena pola huruf Arab yang memiliki kemiripan pada beberapa karakter. Tingkat akurasi pada sistem dipengaruhi oleh jumlah transisi, ukuran normalisasi, jumlah learning rate dan jumlah neuron hidden layer. Akurasi sistem dalam pengenalan pola huruf Arab mencapai 74,3 %.

Kata Kunci : Huruf Arab, Learning Vector Quantization (LVQ), Modified Direction Feature (MDF), Pengenalan Pola

Telkom
University

Abstract

At this time, pattern recognition system has been developed and being used. One of the common pattern recognition that people know is hand-writing recognition. Arabic letter is the letter that Moslem must studied, also has resemblance of pattern at some letters that make pattern recognition of Arabic letter became difficult.

Based on that problems, pattern recognition system that has ability to recognizing hand-writing Arabic letter with output to read the letter that correct by using combination of Modified Direction Feature (MDF) and Learning Vector Quantization (LVQ) method, analysis accuracy from Arabic letter recognition by using MDF and LVQ.

MDF method is technique that get the feature vector from all direction and add together until specific feature vector formed. From feature vector that taken from all direction, can be seen uniqueness from Arabic letter for the input of LVQ that classifying pattern that each of the outputs represent category or class until get the maximum result.

Arabic letter pattern recognition system by using MDF and LVQ make conclusion that MDF and LVQ method can be use in pattern extraction and Arabic letter classification, where pattern feature vector that produced by MDF have a not too high value and can give the uniqueness of pattern, classification from LVQ have a not too high accuracy. This is because Arabic letter's pattern have similarity with another. Accuracy of system influenced by transition, normalisation, learning rate and neuron of hidden layer. Accuracy of the system for Arabic letter recognition reach 74,3 %.

Keywords : Arabic Letter , Learning Vector Quantization (LVQ), Modified Direction Feature (MDF), Pattern Recognition



1. PENDAHULUAN

1.1. LATAR BELAKANG MASALAH

Dalam suatu kalimat, huruf merupakan suatu informasi terkecil yang perlu diidentifikasi dengan baik agar informasi yang ada dalam suatu kalimat dapat lebih dipahami. Huruf memiliki bentuk yang unik dan untuk membedakan antara huruf yang satu dengan yang lain menggunakan karakteristik dari masing-masing huruf tersebut. Pada umumnya proses pengklasifikasian dan pengidentifikasiannya paling sederhana adalah menggunakan penglihatan. Dari sini, timbullah ide bagaimana jika hal tersebut dilakukan oleh proses komputer agar lebih otomatis. Dengan menggunakan pemodelan visual untuk membedakan antar huruf tersebut sehingga untuk pengidentifikasiannya menggunakan pengenalan pola dari huruf tersebut.

Perkembangan penelitian tentang pengenalan pola huruf sudah sangat berkembang, akan tetapi perkembangan pengenalan pola huruf Arab belum sepesat pengenalan huruf lainnya. Hal ini dikarenakan pengenalan huruf Arab memiliki tingkat kesulitan yang lebih tinggi dibandingkan pengenalan huruf lainnya [1]. Padahal, huruf Arab merupakan salah satu huruf yang wajib dipelajari oleh umat muslim di seluruh dunia, dan masing-masing orang berbeda dalam gaya penulisannya. Maka dari itu, diperlukan adanya pengenalan pola huruf Arab yang akan memudahkan orang dalam mempelajari dan mengenali huruf Arab.

Pengenalan pola huruf Arab dilakukan dua tahap, yaitu ekstraksi ciri dan klasifikasi. Untuk melakukan ekstraksi ciri dilakukan dengan metode *Modified Direction Feature* (MDF). MDF mengkombinasikan keunggulan dari teknik *traditional direction feature extraction* dan *transition feature extraction*, yaitu mengkombinasikan ciri arah dan informasi struktur global yang ada pada karakter [4]. Pendekatan yang dilakukan oleh metode ini yaitu dengan melakukan deteksi arah dari segmen-segmen garis pembentuk karakter dan pixel-pixel pembentuk karakter tersebut, untuk kemudian digantikan dengan nilai arah yang sesuai. Akhirnya ciri yang didapatkan dari karakter akan menyimpan lokasi dari pixel dan nilai arahnya. Vector ciri pola yang dihasilkan MDF memiliki nilai yang cukup tinggi dan dapat memberikan keunikan pada pola, selain itu MDF cocok untuk dipasangkan dengan jaringan syaraf tiruan manapun. Ciri yang didapatkan ini nantinya dapat dijadikan input untuk proses pembelajaran jaringan syaraf tiruan maupun proses klasifikasi. Karena hal tersebut, maka MDF digunakan sebagai ekstraksi ciri huruf Arab yang mempunyai banyak lengkungan dalam pola hurufnya.

Proses pengklasifikasian dan pengenalan huruf Arab menggunakan jaringan syaraf tiruan *Learning Vector Quantization* (LVQ). LVQ adalah metode klasifikasi pola yang masing-masing unit keluaran mewakili kategori atau kelas tertentu. LVQ melakukan pembelajaran pada suatu lapisan kompetitif yang akan secara otomatis belajar untuk mengklasifikasikan vektor-vektor input. Jika dua vektor input mendekati sama, maka lapisan kompetitif akan meletakkan kedua vektor input tersebut ke dalam kelas yang sama. Algoritma LVQ bertujuan akhir mencari nilai bobot yang sesuai untuk mengelompokkan vektor-vektor input kedalam kelas tujuan yang telah di inisialisasi pada saat pembentukan Jaringan LVQ. Dalam LVQ membutuhkan sedikit contoh

latihan, lebih cepat dari metode klasifikasi lainnya, dan memproses data yang memiliki *missing value* [2]. Hal tersebut yang menjadikan pertimbangan dalam pemilihan LVQ sebagai pengklasifikasian huruf Arab.

1.2 PERUMUSAN MASALAH

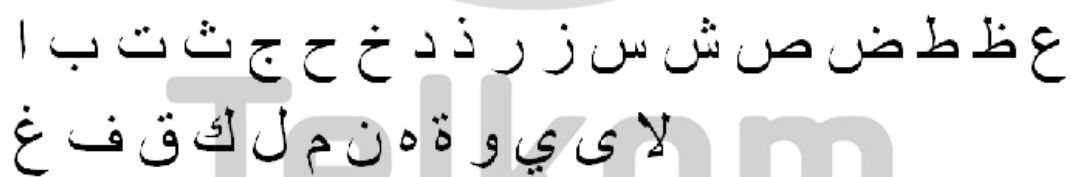
Perumusan masalah dari Tugas Akhir ini berdasarkan latar belakang masalah yang dipaparkan di atas adalah :

- a) Bagaimana mengimplementasikan *Modified Direction Feature* (MDF) dan *JST Learning Vector Quantization* (LVQ).
- b) Bagaimana menghitung performansi akurasi sistem yang menggunakan MDF dan LVQ.
- c) Bagaimana analisis metode MDF dan LVQ pada pengenalan pola huruf Arab berdasarkan parameter-parameter dari metode tersebut.

1.3. BATASAN MASALAH

Batasan masalah dari Tugas Akhir ini adalah :

- a) Pola huruf Arab yang dikenali sistem berupa huruf dasar (tanpa tanda baca dan berdiri sendiri) seperti yang ditunjukkan pada gambar 1.1.
- b) Pola inputan dengan posisi tertentu, tidak menangani tulisan dengan posisi bertumpuk.
- c) Pola inputan untuk proses training dan uji ini berupa huruf Arab dari citra tulisan komputer atau *true type font* (TTF).
- d) Jumlah inputan untuk proses training adalah enam TTF untuk dimodifikasi menjadi berjumlah 15 untuk masing-masing huruf.
- e) Pola inputan untuk proses uji tidak menangani tulisan huruf Arab kaligrafi.
- f) Sebelum masuk dalam proses MDF terlebih dahulu dilakukan beberapa tahap *preprocessing* seperti *Threshold* , *Thinning* , *Crop Edge*, dan Normalisasi.



Gambar 1.1 : Pola Huruf Arab yang akan dikenali sistem

1.4. TUJUAN PENELITIAN

Tujuan dari tugas akhir ini adalah :

- a) Mengimplementasikan kombinasi MDF dan LVQ dalam pengenalan pola huruf Arab.
- b) Menganalisis parameter MDF (jumlah transisi, normalisasi gambar) dan LVQ (jumlah *hidden neuron*, *learning rate*) yang mempengaruhi performansi akurasi sistem dalam pengenalan huruf Arab yang ada di dalam citra.

- c) Menganalisis kombinasi MDF dan LVQ dengan melihat performansi akurasi sistem berdasarkan nilai CER (*Character Error Rate*) mengenai ketepatan pengenalan pola.

1.5. METODOLOGI PENELITIAN

Adapun metodologi yang digunakan dalam penulisan Tugas Akhir ini adalah:

a) Studi Literatur

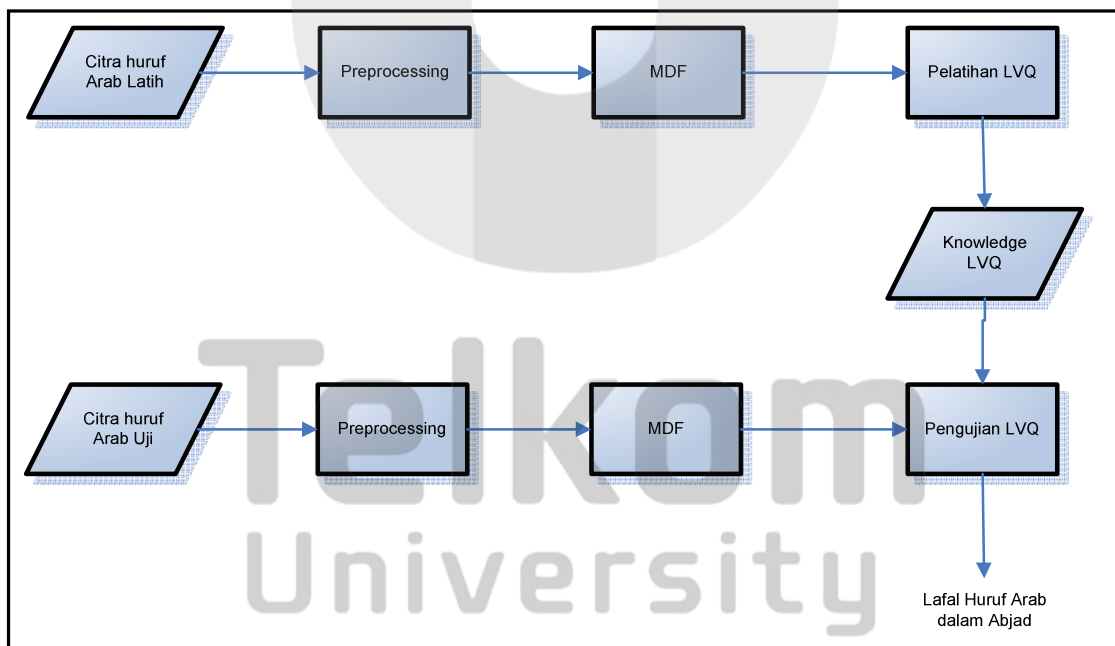
Mengumpulkan informasi dan referensi dari buku, majalah, artikel maupun internet yang akan digunakan sebagai teori dasar penyusunan Tugas Akhir yang berkaitan dengan *Modified Direction Feature* (MDF) dan *Learning Vector Quantization* (LVQ).

b) Analisis Kebutuhan dan Perancangan Sistem

Tahap ini meliputi analisis kebutuhan untuk merancang sistem pengenalan huruf Arab dengan MDF dan LVQ. Desain dari sistem menggunakan pendekatan struktural seperti pada gambar 1.2.

c) Implementasi sistem

Melakukan implementasi pada sistem sesuai dengan analisis kebutuhan dan perancangan yang telah dilakukan pada tahap sebelumnya.



Gambar 1.2 : Flow Diagram Sistem yang Dibangun

d) Analisis Hasil Implementasi

Menganalisis hasil implementasi sistem berkaitan dengan performansi berdasarkan tingkat akurasi dan parameter-parameter yang ada pada MDF dan LVQ yang disesuaikan dengan skenario pengujian.

e) Pembuatan Laporan Tugas Akhir.

Pada tahap akhir, dilakukan pembuatan dokumentasi yang berupa laporan Tugas Akhir



5. PENUTUP

5.1 KESIMPULAN

Dari hasil analisis percobaan yang telah dilakukan pada penelitian ini maka didapatkan beberapa kesimpulan sebagai berikut:

1. Ukuran normalisasi citra inputan memiliki pengaruh yang besar, karena akan mempengaruhi keakuratan ciri yang didapat dari ekstraksi ciri.
2. Jumlah transisi pada MDF sangat berpengaruh besar pada seberapa baik ciri yang dihasilkan, karena tergantung dari pola huruf yang nanti akan di ekstraksi ciri.
3. Pemakaian *hidden layer* juga berpengaruh, dimana semakin besar *hidden layer* semakin bagus tingkat akurasi sistem dalam pengenalan.
4. Learning rate juga berpengaruh ketika dalam peng-update-an bobot pada saat pembelajaran. Nilai learning rate yang tepat akan menghasilkan pembobotan yang sesuai dan akan menghasilkan tingkat akurasi pengenalan yang baik.
5. Tingkat akurasi terbaik dihasilkan sebesar 74,2857 % untuk data uji, dan 76,1904 % untuk data training dimana parameter yang digunakan adalah learning rate 0,0001, *hidden layer* sejumlah 112, jumlah transisi 4, dan ukuran normalisasi 140 x 60.
6. Pengenalan huruf Arab dengan TTF mempunyai tingkat keakuratan yang baik untuk dikenali oleh sistem dibanding dengan tulisan tangan. Karena tulisan tangan untuk masing-masing orang berbeda dalam gaya penulisannya, dan kadang tidak sesuai dengan bentuk baku yang sudah ditetapkan.

5.2 SARAN

Selanjutnya bisa digunakan pre-processing yang lebih tepat untuk diimplementasikan pada kasus pengenalan pola pada huruf Arab. Penghilangan atau pengurangan *noise* juga bisa ditambahkan sebelum dilakukan ekstraksi ciri, agar ciri yang dihasilkan pun akan optimal.

Telkom
University

DAFTAR PUSTAKA

- [1] Sa'adah, Umi., Arif Djunaidy , Aris Tjahyanto , Tri Harsono. 2002. Perangkat Lunak Untuk Aplikasi Pengenalan Karakter Arab, Available on: <http://www.ies.eepis-its.edu/prosiding/abstrakdetail.php?id=254> (5 agustus 2009)
- [2] WEKA Classification Algorithms. Available on: <http://weka.classalgos.sourceforge.net/> (5 agustus 2009)
- [3] Gader, Paul D., Magdi Mohamed, Jung-Hsien Chiang.1997. *Handwritten Word Recognition with Character and Inter-Character Neural Networks*.
- [4] Blumenstein, M., Liu, X. Y., & Verma, B.2004. *An Investigation of the Modified Direction Feature for Cursive Character Recognition*
- [5] Arymurthi, Aniati Murni. 1992. *Pengantar Pengolahan Citra*. Jakarta: PT. ElexMedia Komputindo
- [6] Bullinaria, John. A.2004. *Learning Vector Quantization (LVQ)*.
- [7] S., Dijas Schwartz. 2005. *Optical Character Recognition (OCR)*. Universitas Gadjah Mada
- [8] Suyanto.2007. *Artificial Intelligence*. Informatika Bandung
- [9] Hermanto, I Gede Rudy., Tjokorda Agung Budi W, Retno Novi D. 2008. *Analisis Dan Implementasi Pengenalan Huruf Bali Menggunakan Modified Direction Feature Dan Jaringan Syaraf Tiruan*. Institut Teknologi Telkom.
- [10] Sari, Devia Purnama., Retno Novi D, Rimba Whidiana. 2008. *Implementasi Pendekatan Jaringan Saraf Tiruan ART-2 Pada Pengenalan Pola Huruf Arab*. Institut Teknologi Telkom.