

PERANCANGAN DAN ANALISIS SISTEM PENGENALAN KATA AKSARA SUNDA MENGUNAKAN METODE LEARNING VECTOR QUANTIZATION BERBASIS PENGOLAHAN CITRA

Designing and Analysis Of Recognition Sundanese Character Word System Using Thresholding Method Based On Image Processing

Adri Achmad Farhan¹, Ratri Dwi Atmaja, ST.,M.T.², Suci Aulia, S.T., M.T.³

^{1,2} Prodi S1 Teknik Telekomunikasi, Fakultas Teknik Elektro, Universitas Telkom

³ Prodi D3 Teknik Telekomunikasi, Fakultas Ilmu Terapan, Universitas Telkom

¹adrifarhan.af@gmail.com, ²ratriidwiatmaja@tess.telkomuniversity.ac.id, ³sucia@tass.telkomuniversity.ac.id

Abstrak

Aksara Sunda Merupakan salah satu sejarah di Indonesia dalam hal aksara atau penulisan. Aksara sunda Aksara Sunda berjumlah 32 buah. Terdiri atas 7 aksara vokal atau aksara swara dan 23 aksara konsonan atau aksara *ngalagena*. Pengenalan bahasa merupakan salah satu elemen penting dari sebuah komunikasi, terlebih lagi huruf atau tulisan tangan. Kedua hal tersebut merupakan salah satu bidang pengenalan pola yang telah memberikan kontribusi yang cukup besar bagi kemajuan perkembangan suatu teknologi.

Penelitian ini merancang sistem yang dapat mendeteksi atau mengenali aksara sunda berbasis pengolahan citra, menggunakan metode *Learning Vector Quantization* (LVQ). pada penelitian kali ini menggunakan ekstraksi ciri yang digunakan adalah penjumlahan biner (horizontal dan vertical), DCT, dan DFT.

Hasil dari pengujian pada 30 kata aksara sunda terhadap 4 ekstraksi ciri yang diujikan yaitu didapatkan bahwa tingkat akurasi tertinggi sebesar 83.33% pada ekstraksi ciri DCT, sedangkan pada ekstraksi ciri Penjumlahan Vertikal didapatkan akurasi sebesar 76.67%, lalu sebesar 73.33% pada ekstraksi ciri Penjumlahan horizontal, dan pada ekstraksi ciri DFT didapatkan akurasi sebesar 60%.

Kata Kunci: Aksara Sunda, Pengolahan Citra, *Thresholding*, *Learning Vector Quantization* (LVQ).

Abstract

Sundanese script is a historical one in Indonesia in terms of script or writing. Literacy Sundanese Sundanese script totaling 32 pieces. Consisting of 7 characters or characters swara script and 23 consonants script or script Ngalagena. The introduction of language is one of the essential elements of a communication, even more letters or handwritten. Secondly it is one of the fields of pattern recognition that has contributed significantly to the advancement of the development of a technology.

This study design systems that can detect or recognize characters Sunda-based image processing, using Learning Vector Quantization (LVQ). in the present research using feature extraction that is used is the summation of binary (horizontal and vertical), DCT and DFT.

The results of testing at 30 words script Sunda terhadap 4 extraction characteristic tested is found that the highest level of accuracy of 83.33% on the extraction characteristics of DCT, while the extraction characteristics Penjumlahan Vertical obtained an accuracy of 76.67%, and amounted to 73.33% on the extraction characteristics Penjumlahan horizontal, and the extraction characteristics of DFT obtained an accuracy of 60%.

Key Words: *Sundanese script, Image Processing, thresholding, Learning Vector Quantization (LVQ).*

1. Latar Belakang

Indonesia dikenal sebagai negara kepulauan dan memiliki beribu-ribu warisan budaya yang tersebar di seluruh penjuru nusantara. Diantara banyaknya warisan budaya tersebut adalah karakter atau tulisan asli berbagai daerah yang termasuk di dalam kategori Aksara Nusantara. Secara garis besar aksara nusantara dapat dikelompokkan menjadi beberapa kelompok besar diantaranya Aksara Hanacaraka (ada di Jawa, Sunda dan Bali), Aksara Batak (ada di Toba, Simalungun, Karo), dan Aksara Ulu (ada di Kerinci, Rejang, Lampung, Serawi) [1] [2]. Aksara sunda merupakan salah satu peninggalan sejarah asli masyarakat sunda, sejak abad ke 5, aksara sunda sunda menjadi bahasa penulisan pada saat itu. Namun semakin majunya jaman, semakin pudar pula eksistensi dari aksara sunda itu sendiri. Dimana anak anak dari daerah sunda sendiri pun banyak yang tidak paham atau mengerti tentang bentuk dan arti dari aksara sunda itu sendiri.

Pada penelitian sebelumnya [3] telah dilakukan penelitian tentang pengenalan kata dalam aksara sunda menggunakan metode deteksi tepi dan Learning Vector Quantization (LVQ) berbasis pengolahan citra pada android. Dalam penelitian ini dirancang sebuah sistem yang dapat mendeteksi atau mengenali aksara sunda berbasis pengolahan citra. Masukan dalam aplikasi ini berupa aksara sunda yang ditulis tangan yang akan disegmentasi dan akan menghasilkan nilai-nilai piksel yang kemudian dijadikan ciri suatu aksara sunda. Kemudian ciri tersebut akan diklasifikasi menggunakan metode Learning Vector Quantization (LVQ). pada penelitian kali ini menggunakan segmentasi thresholding dan ekstraksi ciri yang digunakan adalah penjumlahan biner, DCT, dan DFT.

2. Aksara Sunda[3]

Aksara Sunda berjumlah 32 buah. Terdiri atas 7 aksara vokal atau aksara swara (a – é – i – o – u – e – eu) dan 23 aksara konsonan atau aksara *ngalagena* (ka – ga – nga, ca – ja – nya, ta – da – na, pa – ba – ma, ya – ra – la, wa – sa – ha, fa – va – qa – xa – za). Aksara-aksara “fa”, “va”, “qa”, “xa”, dan juga “za” merupakan aksara-aksara yang relatif baru.

Aksara ini diciptakan untuk mengonversi beberapa bunyi aksara Latin. Dari segi bentuknya, aksara Sunda ini memiliki bentuk persegi yang memperlihatkan ketajaman yang mencolok, dan hanya sebagian saja yang memiliki bentuk bundar.

3. Citra Digital[6]

Citra digital merupakan representatif dari citra yang diambil oleh mesin dengan bentuk pendekatan berdasarkan sampling dan kuantisasi. Citra terdiri dari 3 jenis, yaitu binary image, grayscale, dan color image. Binary image merupakan bentuk citra dimana dalam setiap pixelnya hanya mempunyai dua kemungkinan nilai yaitu 0 dan 1 (berukuran 1 bit) atau dengan kata lain citra ini hanya mempunyai dua buah warna. Sedangkan color image atau yang sering disebut sebagai citra RGB merupakan citra yang terdiri dari kombinasi tiga komponen warna yaitu R (Red), G (Green) dan B (Blue). Jenis citra ini biasanya berukuran 24 atau 32 bit. Dan citra grayscale merupakan jenis citra dimana setiap pixelnya direpresentasikan dengan satu nilai yang menunjukkan nilai warna dari hitam ke putih (gray level).

4. Pengolahan Citra Digital[8]

Pengolahan citra digital dapat dikelompokkan dalam dua jenis, yaitu:

- Memperbaiki kualitas suatu gambar sehingga dapat lebih mudah diinterpretasi oleh mata manusia.
- Mengolah informasi yang terdapat pada suatu gambar untuk keperluan pengenalan objek secara otomatis.

Pengenalan pola (pattern recognition) pada umumnya bertujuan untuk mengenali suatu objek dengan cara mengekstrak informasi penting yang dapat menjadi ciri khusus bagi objek tersebut. Pengenalan pola dan pengolahan citra dapat dikombinasikan sehingga membentuk suatu sistem yang dapat memproses citra masukan sehingga citra tersebut dapat dikenali polanya. Proses ini disebut pengenalan citra atau image recognition.

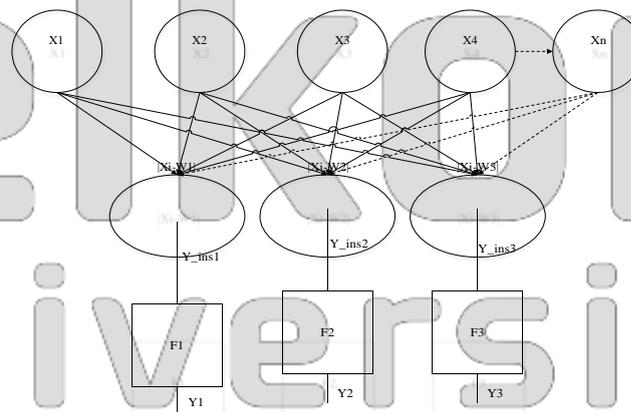
5. Metode Thresholding[8]

Thresholding merupakan salah satu metode segmentasi citra di mana prosesnya didasarkan pada perbedaan derajat keabuan citra. Dalam proses ini dibutuhkan suatu nilai batas yang disebut nilai threshold. Nilai intensitas citra yang lebih dari atau sama dengan nilai threshold akan diubah menjadi putih (1) sedangkan nilai intensitas citra yang kurang dari nilai threshold akan diubah menjadi hitam (0). Sehingga keluaran dari hasil thresholding adalah berupa citra biner.

6. Leringan Vector Quantization

Learning Vector Quantization (LVQ) adalah sebuah metode klasifikasi dimana setiap unit output mempresentasikan sebuah kelas. LVQ digunakan untuk pengelompokan dimana jumlah kelompok sudah ditentukan arsitekturnya (target/kelas sudah ditentukan). LVQ adalah salah satu jaringan syaraf tiruan yang merupakan algoritma pembelajaran kompetitif dari algoritma Kohonen Self-Organizing map (SOM). Tujuan dari algoritma ini adalah untuk mendekati distribusi kelas vektor untuk meminimalkan kesalahan dalam pengklasifikasian. Algoritma diusulkan oleh Kohonen pada tahun 1986 sebagai perbaikan dari Vector Quantization.

Model pembelajaran LVQ dilatih secara signifikan agar lebih cepat dibandingkan algoritma lain seperti Back Propagation Neural Network. Hal ini dapat meringkas atau mengurangi dataset besar untuk sejumlah kecil vektor. LVQ melakukan pembelajaran pada lapisan kompetitif yang terawasi. Lapisan kompetitif akan secara otomatis belajar untuk mengklasifikasikan vektorvektor input. Kelas-kelas yang idapat sebagai hasil dari lapisan kompetitif ini hanya tergantung pada jarak antara vektor-vektor input. Jika vektor input mendekati sama maka lapisan kompetitif akan mengklasifikasikan kedua vektor input tersebut kedalam kelas yang sama. [4]



Gambar 1 Jaringan Syaraf LVQ

7. Ekstraksi Ciri

Ekstraksi ciri merupakan cara yang dilakukan untuk mendapatkan ciri dari sebuah citra. Proses ini merupakan salah satu tahap yang penting dalam mendeteksi citra. Melalui ekstraksi ciri akan diperoleh informasi-informasi penting dari citra aksara sunda yang dapat membedakan satu kata dengan kata lainnya. Pada penelitian ini menggunakan 4 metoda ekstraksi ciri yaitu

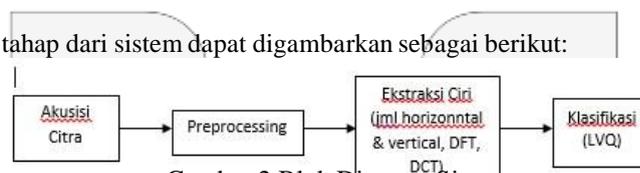
- a. Penjumlahan horizontal
Penjumlahan horizontal merupakan penjumlahan biner secara horizontal dan menjadikannya sebagai sebuah ciri
- b. Penjumlahan vertical
Penjumlahan vertikal merupakan penjumlahan biner secara horizontal dan menjadikannya sebagai sebuah ciri
- c. DCT(Discrete Cosine Transform)
Merupakan teknik yang digunakan untuk dapat melakukan konversi sinyal kedalam komponen frekuensi pembentukannya, dengan cara memperhitungkan nilai riil dari hasil transformasinya.
- d. DFT (Discrete Fourier Transform)
Memiliki basis sinyal dan merupakan bentuk kompleks. Sehingga memiliki frekuensi yang dihasilkan juga memiliki bentuk kompleks. Dengan demikian dapat terlihat adanya bagian riil dan imajiner, dan dapat juga dilihat hasil transformasi dalam bentuk sinyal yang absolute yang biasa disebut magnitude respon frekuensi dan magnitude respon fase.

8. Blok Sistem Pengenal Pola Aksara Sunda

Secara umum, pemodelan sistem penerjemah aksara sunda pada penelitian ini terbagi menjadi kedalam beberapa tahapan, yaitu:

- a. Tahap akuisisi citra
- b. Tahap preprocessing
- c. Tahap ekstraksi ciri
- d. Tahap klasifikasi

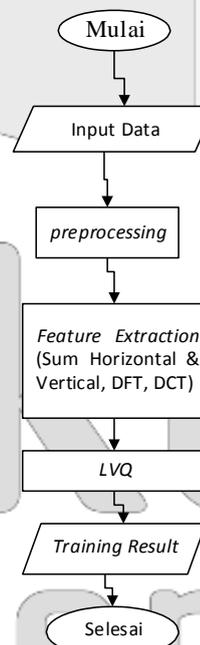
Secara keseluruhan tahap – tahap dari sistem dapat digambarkan sebagai berikut:



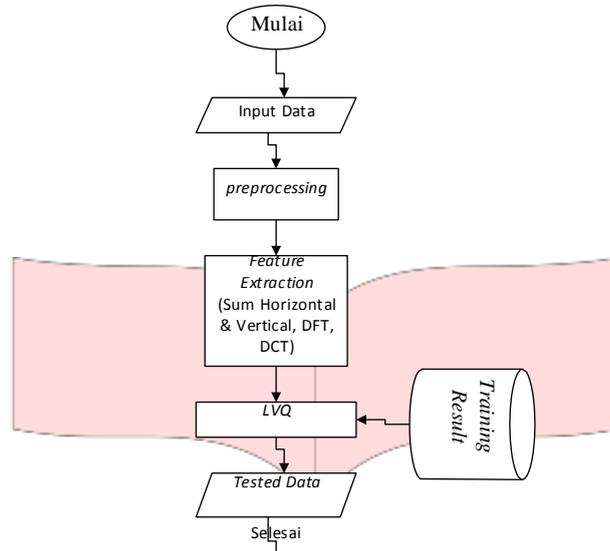
Gambar 2 Blok Diagram Sistem

9. Proses Latih dan Uji

Sistem yang akan dibuat dalam penelitian ini terdiri dari 2 macam proses, yang pertama adalah proses training yang merupakan proses pembentukan database yang menyimpan nilai ciri dari tiap citra yang dijadikan sebagai referensi bagi citra uji nantinya. Lalu yang berikutnya adalah proses testing merupakan proses percobaan database yang telah dibuat pada proses sebelumnya.



Gambar 3 Flowchart Training Data[8]



Gambar 4 Flowchart Tested Data[8]

10. Proses Akuisisi Citra

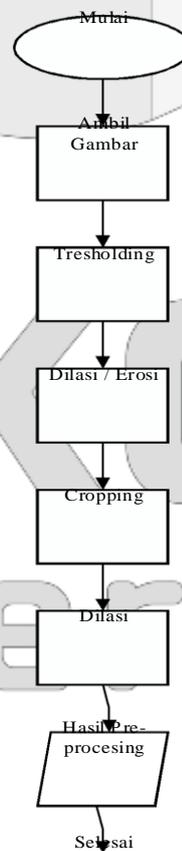
Akuisisi citra merupakan proses pengambilan citra gambar yang akan digunakan sistem. Dalam tugas akhir ini, akuisisi citra didapat dari pengambilan gambar yang akan di proses pada matlab. Citra yang digunakan berupa kata dalam aksara sunda yang ditulis tangan pada kertas HVS. Data citra yang telah diambil kamera handphone lalu dilakukan crop dengan ukuran 150x200 pixel. Data gambar yang telah dikumpulkan dilakukan pengetesan terhadap masing-masing sampel data baik sebagai data latih maupun data uji.

10.1. Input

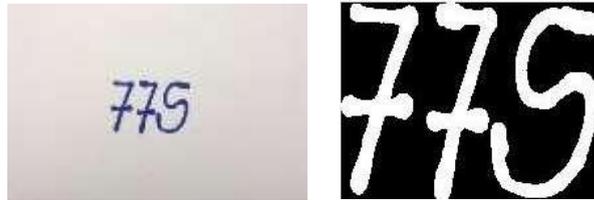
Proses input pertama kali dilakukan sebagai proses masukan citra hasil dari pengambilan gambar tulisan tangan yang nanti akan dilakukan pengenalan huruf .File masukan berupa *.jpg.

10.2. Pre-Processing

Preprocessing merupakan proses yang dilakukan pada suatu citra digital sebelum dilakukan pemrosesan citra selanjutnya. Tujuan dari preprocessing adalah untuk meningkatkan kualitas dari citra masukan yang akan didapat.



Gambar 5 Diagram Alir Proses Pre-Processing



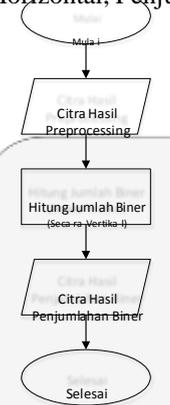
(a) (b)
Gambar 6 (a) Citra Asli (b) Citra Setelah Preprocessing

10.3. Ekstraksi Ciri

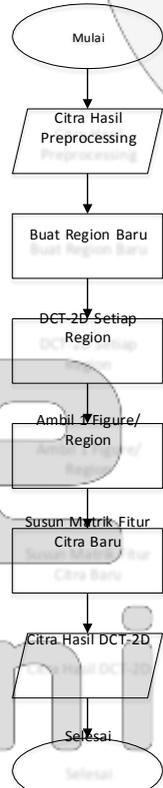
Tahap Selanjutnya adalah melakukan ekstraksi ciri Terhadap citra yang telah dilakukan preprocessing. Ekstraksi ciri merupakan proses untuk mendapatkan ciri atau fitur dari gambar input yang akan dilakukan proses pengklasifikasian dengan menggunakan metode Learning Vector Quantization (LVQ). Setelah citra di normalisasi ke ukuran yang telah di tentukan, maka selanjutnya akan dilakukan pencirian. Ada 4 Metode yang digunakan untuk ekstraksi ciri dalam Tugas Akhir ini yaitu Penjumlahan Horizontal, Penjumlahan Vertikal, DCT dan DFT .



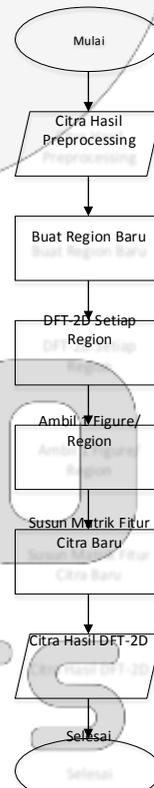
Gambar 7 Diagram Alir Jumlah Horizontal



Gambar 8 Diagram Alir Jumlah Vertikal



Gambar 9 Diagram Alir DCT-2D[11]

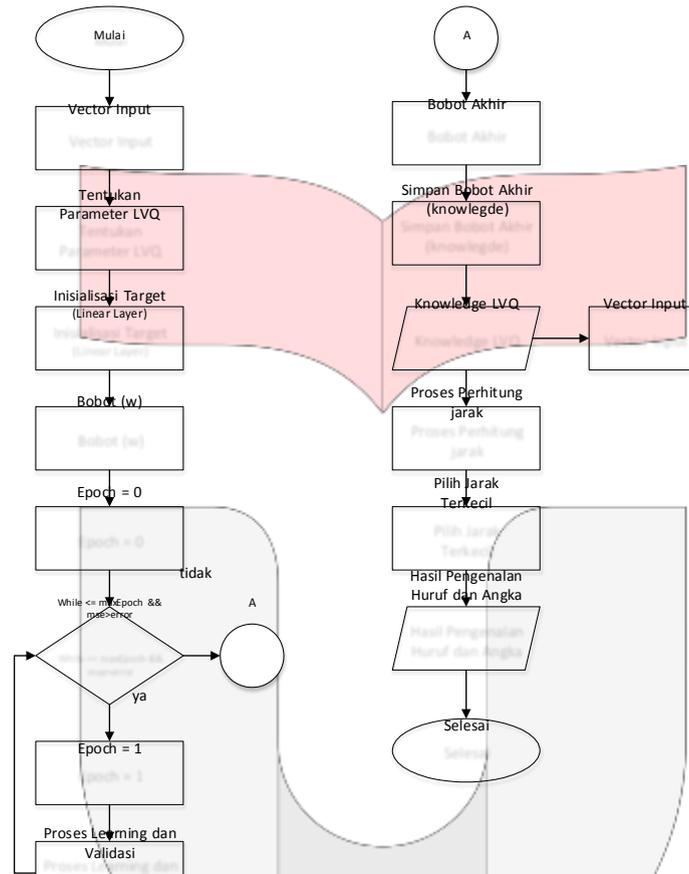


Gambar 10 Diagram Alir DFT-2D[11]

10.4. Klasifikasi

Pada tugas akhir ini proses klasifikasi dibagi menjadi 2 proses yaitu proses training dan proses testing. Metode yang digunakan dalam tugas akhir ini adalah metode Learning Vector Quantization (LVQ). Langkah pertama adalah inialisasi pada proses klasifikasi, dilakukan untuk menentukan bobot awal suatu kelas. Setelah dilakukan inialisasi sistem akan membandingkan bobot awal tiap kelas tersebut dengan nilai input dan

menghitung jarak terdekat. Jika terdapat nilai input yang mendekati sama maka lapisan kompetitif akan meletakkan nilai tersebut ke dalam kelas yang sama. Pada proses klasifikasi dilakukan perubahan pada ukuran layer yaitu ukuran 50, 100, 150, 200, 250..



Gambar 11 Diagram Alir Learning Vector Quantization[12]

11. Pengujian Sistem

Pengujian sistem dilakukan untuk menguji sistem yang dirancang dapat berjalan dengan baik atau tidak. Pengujian ini menggunakan masukan berupa data uji dan data latih tulisan kata dalam aksara sunda. Penulisan data dilakukan oleh responden yang memahami tulisan aksara sunda.

12. Analisis Performansi Sistem

Parameter yang digunakan dalam menentukan keakuratan sistem dengan menguji pengaruh jenis perubahan pixel, perbandingan antara 3 ekstrasi ciri, serta perubahan layer pada klasifikasi LVQ yang digunakan. Akurasi sistem dihitung dengan membandingkan jumlah data yang diidentifikasi dengan benar dengan jumlah data uji total yang digunakan. Berikut adalah persamaan dalam menentukan keakuratan sistem.
 $Akurasi = \frac{\text{Jumlah Data Benar}}{\text{Jumlah Total Data}} \times 100\%$

13. Pengaruh Ekstraksi Ciri Terhadap Proses Latih

13.1. Pengujian Ekstraksi Ciri Jumlah Horizontal Terhadap Beberapa Layer

Dari hasil analisis yang dilakukan bahwa didapatkan tingkat akurasi paling tinggi sebesar 73,33% pada layer 50, sebesar 70% pada layer 100, sebesar 66.67% pada layer 150, sebesar 70% pada layer 200, sebesar 66.67% pada layer 250. Dapat dilihat pada grafik dibawah.

Tingkat akurasi rata-rata pengaruh topologi JST terhadap proses pelatihan dapat dilihat pada diagram berikut.



Gambar 12 Tingkat Akurasi Tertinggi Pada Ekstraksi Ciri Jumlah Horizontal

13.2. Pengujian Ekstraksi Ciri Jumlah Vertikal Terhadap Beberapa Layer

Dari hasil analisis yang dilakukan bahwa didapatkan tingkat akurasi paling tinggi sebesar 66.67% pada layer 50, sebesar 73.33% pada layer 100, sebesar 66.67% pada layer 150, sebesar 63.33% pada layer 200, sebesar 76.67% pada layer 250. Dapat dilihat pada grafik dibawah.



Gambar 13 Tingkat Akurasi Tertinggi Pada Ekstraksi Ciri Jumlah Vretikal

13.3. Pengujian Ekstraksi Ciri DFT Terhadap Beberapa Layer

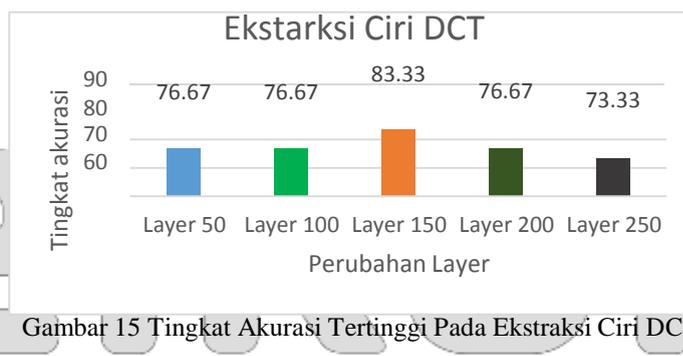
Dari hasil analisis yang dilakukan bahwa didapatkan tingkat akurasi paling tinggi sebesar 56.67% pada layer 50, sebesar 53.33% pada layer 100, sebesar 60% pada layer 150, sebesar 60% pada layer 200, sebesar 60% pada layer 250. Dapat dilihat pada grafik dibawah.



Gambar 14 Tingkat Akurasi Tertinggi Pada Ekstraksi Ciri DFT

13.4. Pengujian Ekstraksi Ciri DCT Terhadap Beberapa Layer

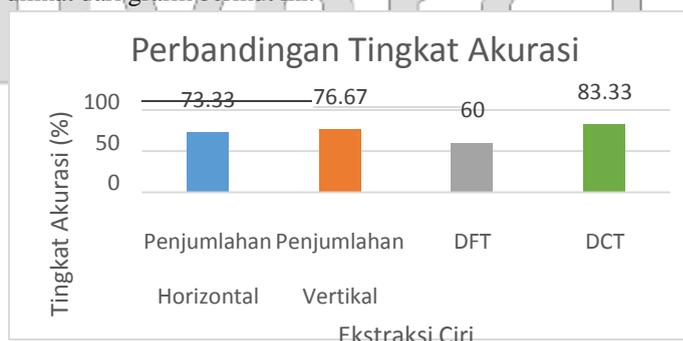
Dari hasil analisis yang dilakukan bahwa didapatkan tingkat akurasi paling tinggi sebesar 76.67% pada layer 50, sebesar 76.67% pada layer 100, sebesar 83.33% pada layer 150, sebesar 76.67% pada layer 200, dan sebesar 73.33% pada layer 250. Dapat dilihat pada grafik dibawah.



Gambar 15 Tingkat Akurasi Tertinggi Pada Ekstraksi Ciri DCT

13.5. Perbandingan tingkat akurasi antara ekstraksi ciri

Dari hasil analisis, dapat diketahui bahwa pada ekstraksi ciri DCT mempunyai hasil akurasi paling tinggi dibandingkan ekstraksi ciri lainnya, dapat dilihat dari grafik berikut ini.



Tabel 16 Perbandingan Tingkat Akurasi Setiap Ekstraksi Ciri

14. Kesimpulan

Diketahui Berdasarkan hasil pengujian dan analisis pada sistem pengenalan kata aksara sunda pada pengujian kali ini didapatkan hasil setelah dilakukan pengujian data yang menggunakan metode ekstraksi ciri penjumlahan *horizontal*, penjumlahan *vertical*, *Discrete Fourier Transform* (DFT) dan *Discrete Cosine Transform* (DCT) serta metode klasifikasi menggunakan Jaringan Syaraf Tiruan *Learning Vector Quantization*, maka didapatkan kesimpulan sebagai berikut:

1. Metode ekstraksi ciri penjumlahan *horizontal*, penjumlahan *vertical*, *Discrete Fourier Transform* (DFT) dan *Discrete Cosine Transform* (DCT) dan metode klasifikasi Jaringan Syaraf Tiruan *Learning Vector Quantization Map* (LVQ) layer dapat digunakan untuk mengenal pola kata aksara sunda.
2. Metode ekstraksi ciri *Discrete Cosine Transform* (DCT) merupakan metode yang baik pada *Learning Vector Quantization Map* (LVQ) karena dapat menghasilkan tingkat akurasi yang paling tinggi dengan tingkat akurasi sebesar 83.33%. sedangkan pada ekstraksi ciri penjumlahan Horizontal didapatkan tingkat akurasi sebesar 73.33%. pada ekstraksi ciri penjumlahan vertical didapatkan tingkat akurasi sebesar 76.67%. dan pada ekstraksi ciri DFT akurasi didapatkan sebesar 60%.
3. Jumlah hidden layer pada klasifikasi mempengaruhi tingkat akurasi. pada ekstraksi ciri dct tingkat akurasi paling tinggi didapatkan dengan layer 150, pada penjumlahan Horizontal didapatkan tingkat akurasi tertinggi dengan layer 50. pada ekstraksi ciri penjumlahan vertical didapatkan tingkat akurasi tertinggi dengan layer 250. dan pada ekstraksi ciri DFT akurasi tertinggi didapatkan dengan layer 150,200,250.
4. Perubahan ukuran pixel tidak memberikan perubahan tingkat akurasi yang stabil. Peningkatan jumlah data menyebabkan penurunan tingkat akurasi. Pada ekstraksi ciri DCT dengan 40 data tingkat akurasi tertinggi sebesar 72.5% dan dengan 50 tingkat akurasi tertinggi sebesar 54%.

15. Saran

Sebagai perbaikan untuk Tugas Akhir ini dapat dilakukan pengembangan terhadap sistem, antara lain seperti:

1. Melakukan percobaan dengan ekstraksi ciri yang lain agar didapatkan tingkat akurasi yang lebih baik lagi.
2. Melakukan percobaan dengan klasifikasi yang lain agar didapatkan hasil yang lebih baik.
3. Dilakukan percobaan dengan kata yang lebih banyak atau dalam bentuk kalimat
4. Dapat diimplementasikan dalam bentuk software baik pembelajaran atau berupa permainan agar dapat mempermudah dalam mempelajari aksara sunda

16. Daftar Pustaka

- [1] N. D. Kertasari, N. Haswanto and P. Sunarto, Tipografi Adaptasi Karakter Aksara Batak Toba Dalam Huruf Latin, 2009.
- [2] U. Kozok, Sejarah Perkembangan Tulisan Batak Berikut Pedoman Menulis Aksara Batak dan Cap Si Singamangaraja XII, Kepustakaan Populer Gramedia, 2009.
- [3] T. U. A. Sunda, Direktori Aksara Sunda Untuk Unicode, Bandung: Dinas Pendidikan Provinsi Jawa Barat, 2008.
- [4] A. W. Witoelar, M. Biehl and B. Hammer, "Learning Vector Quantization : generalization ability and dynamics of competing prototypes," in Proc. 6th Int. Work. Self-Organizing Maps, 2007.
- [5] N. A. Putri, G. Budiman and Y. S. Haryani, "ANALISIS DAN IMPLEMENTASI IDENTIFIKASI KEPERIBADIAN MELALUI," 2015.
- [6] D. Putra, Pengolahan Citra Digital, Yogyakarta: Andi, 2010.
- [7] M. E. S. N, G. Budiman and Y. S. Haryani, "IMPLEMENTASI METODE SOBEL EDGE DAN TEMPLATE MATCHING," 2015.
- [8] D. D. Lestari, PERANCANGAN PENGENAL KATA DALAM AKSARA SUNDA MENGGUNAKAN METODE DETEKSITEPI DAN LVQ BERBASIS PENGOLAHAN CITRA PADA ANDROID, Bandung: universitas telkom, 2015.
- [9] M. R. Kumaseh, L. Latumakulita and N. Nainggolan, "SEGMENTASI CITRA DIGITAL IKAN MENGGUNAKAN".
- [10] Sutojo, T., dkk, 2011, Kecerdasan Buatan Yogyakarta, ANDI
- [11] A. Ilham, DESAIN DAN REALISASI SISTEM PENDETEKSI WAJAH UNTUK ABSENSI KARYAWAN DENGAN METODE 2D-DCT DAN PCA MENGGUNAKAN WEBCAM, Bandung: universitas telkom, 2012
- [12] Z. Alfian, IMPLEMENTASI PEMBACAAN HURUF HIJAIYAH DAN KARAKTERISTIK DAN KARAKTER ANGKA ARAB DENGAN MENGGUNAKAN JARINGAN SYARAF TIRUAN LVQ (LEARNING VECTOR QUANTIZATION), Bandung: universitas Telkom, 2013