



Institut Sains dan Teknologi Terpadu Surabaya

Jalan Ngagel Jaya Tengah 73 – 77, Surabaya 60284 , Indonesia
Telp. (031) 5027920 Fax. (031) 5041509

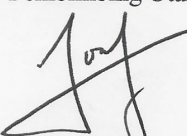
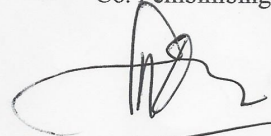
PROPOSAL TESIS

Periode Bulan: November Tahun: 2019
Semester Gasal / Genap *) Tahun ajaran 2019 / 2020

Program Studi : S2 TEKNOLOGI INFORMASI
Nama Mahasiswa : MIFTAH FARID
NRP Mahasiswa : 215210499
Judul Tesis :
KONVERSI CARAKAN MADURA MENJADI HURUF LATIN MENGGUNAKAN ARTIFICIAL
NEURAL NETWORK DAN HIDDEN MARKOV MODEL ✓

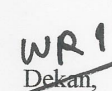
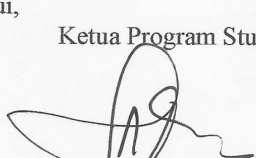


Jenis Tesis : ☐ Kontribusi Keilmuan ☒ Kontribusi Pemanfaatan
Pembimbing Utama : JOAN SANTOSO, S.Kom, M.Kom ✓
Co. Pembimbing : DR. Ir. ENDANG SETYATI, MT ✓
Jumlah SKS **SUDAH LULUS** : 33 SKS IPK : 3,27

Mengetahui,
Pembimbing Utama Co. Pembimbing Surabaya, 11 Februari 2020
  Pemohon
(JOAN SANTOSO, S.Kom, M.Kom) (DR. Ir. ENDANG SETYATI, MT) (MIFTAH FARID)

Catatan Tambahan:

| |
|--|
| |
| |
| |
| |
| |
| |
| |
| |
| |

Surabaya, 11 Februari 2020
Mengetahui,
Dekan, Ketua Program Studi,
 
(GUNAWAN) (Endang Setyati)



Institut Sains dan Teknologi Terpadu Surabaya

Jalan Ngagel Jaya Tengah 73 – 77, Surabaya 60284 , Indonesia

Telp. (031) 5027920 Fax. (031) 5041509

CHECKLIST PROPOSAL TESIS

Dosen pembimbing wajib memberikan check (√) untuk tiap point yang telah dipenuhi.

Ketentuan umum yang harus dipenuhi:

- ☒ Hipotesa tesis bukan merupakan fakta melainkan hasil penalaran yang jelas dan terukur, sedapat mungkin dapat diukur secara kuantitatif.
- ☒ Tinjauan Pustaka minimal 3 (tiga) paper yang menunjang dan/atau mendasari penelitian (diusahakan paper internasional), masing-masing tinjauan pustaka dibahas sepanjang 1(satu) halaman di proposal.
- ☒ Blok diagram proses beserta spesifikasi input-output sudah dijelaskan.
- ☒ Ruang lingkup minimal 2 (dua) halaman dan berisi batasan-batasan penelitian yang ditulis dengan jelas dan bukan merupakan teori-teori dasar.
- ☒ Target keberhasilan tesis sudah dijelaskan.
- ☒ Daftar pustaka minimal 5 paper, termasuk 3 judul paper di tinjauan pustaka.

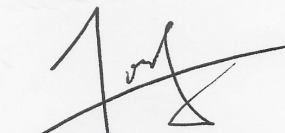
Untuk kontribusi keilmuan:


- ☐ Dataset yang akan digunakan sudah dijelaskan spesifikasi dan jumlah datanya.
- ☐ Peningkatan dan/atau perbaikan yang akan dilakukan sudah dijelaskan .

Untuk kontribusi pemanfaatan:

- ☒ Spesifikasi Institusi tempat penelitian ini dimanfaatkan harus ditulis dengan jelas. (jumlah karyawan, jumlah cabang, jumlah komputer, jumlah jaringan, dan lain-lain tergantung tujuan tesis).
- ☒ Jumlah dan spesifikasi data yang akan digunakan dalam uji coba keberhasilan pemanfaatan tesis sudah jelas.

Surabaya, 11 Februari 2020


JOAN SANTOSO, S.Kom, M.Kom
Dosen Pembimbing


MIFTAH FARID
NRP: 215210499



SEKOLAH TINGGI TEKNIK SURABAYA

Status : **DISAMAKAN** SK. MENDIKBUD REPUBLIK INDONESIA
TERAKREDITASI Badan Akreditasi Nasional Perguruan Tinggi
Jl. Ngagel Jaya Tengah 73-77, Surabaya 60284, P.O. Box 1398, Indonesia
Telp. (031) 5027920 (Hunting) ; Fax. (031) 5041509
Homepage : <http://www.stts.edu>

Hasil Keputusan Sidang Proposal Periode NOVEMBER 2019

Informasi Tugas Akhir/Tesis

NRP : 215210499

Nama : MIFTAH FARID

KONVERSI AKSARA GHAJANG DAN PANGANGGUY
SOWARA CARAKAN Judul : MADURA MENJADI HURUF LATIN
MENGUNAKAN ARTIFICIAL NEURAL
NETWORK DAN HIDDEN MARKOV MODEL

Judul Baru : Konversi Carakan Madura Menjadi Huruf Latin Menggunakan
Artificial Neural Network dan Hidden Markov Model

Pembimbing : Joan Santoso,
S.Kom., M.Kom. Co-Pembimbing :
Endang Setyati, Hj., Dr., Ir., M.T.

Informasi Periode Tugas Akhir/Tesis

Tanggal Rapat : 13 November 2019

Tanggal Cetak : 25 January 2020, 07:22

Review

Reviewer 1 : Hendrawan Armanto,
S.Kom., M.Kom. Status :

DIPERBAIKI

Pesan : 1. Pastikan ini adalah handwriting recognition karena contoh input
adalah tulisan tangan siswa
2. Arahkan ujicoba agar dapat mengenali sebuah kalimat yang
ditulis siswa, bukan hanya mengenali 1 set aksara.
3. Pada perancangan sistem, beri contoh proses nya mulai dari
preprocessing, segmentasi, ekstraksi, dan klasifikasi

Tanggapan :

1. Pengolahan citra yang digunakan berupa handwriting recognition (penjelasan terdapat pada bagian latar belakang).

2. Sesuai hasil rapat, *input* berupa kata carakan madura yang ditulis siswa (penjelasan terdapat pada bagian *input* dan *output* sistem)
3. Contoh proses mulai dari preprocessing sampai klasifikasi sudah dicantumkan pada bagian perancangan sistem.

Reviewer 2 : Yuliana Melita Pranoto,

S.Kom., M.Kom. Status :

DIPERBAIKI

Pesan : 1. Output seharusnya berupa kata/kalimat, tidak sekedar suku kata saja. 2.

Perhatikan segmentasi ketika input saling bertumpuk dibagian atas/bawah. 3. Fitur tekstur untuk apa? Hubungannya dengan bentuk kotak/segitiga? 4. Hati-hati dengan target penelitian yang dituliskan pada hipotesa.

Tanggapan :

1. Sesuai hasil rapat, *input* berupa kata yang ditulis siswa dan *output* berupa kata huruf latin (penjelasan terdapat pada bagian *input* dan *output* sistem).
2. Segmentasi disesuaikan dengan input yang hanya menggunakan 1 kata carakan madura
3. Sesuai petunjuk dari penguji, fitur tekstur tidak jadi digunakan karena tidak ada kaitannya
4. Sesuai petunjuk dari pembimbing, hipotesa diturunkan menjadi 70% (penjelasan terdapat pada bagian hipotesis)

Reviewer 3 : Reddy Alexandro Harianto,

S.Kom., M. Kom. Status : **DITERIMA**

Pesan :

Tanggapan :

Reviewer 4 : Eka Rahayu Setyaningsih,

S.Kom., M.Kom. Status : **DIPERBAIKI**

Pesan : 1. Sebutkan dengan lebih terperinci preprocessing apa saja yang akan dilakukan. Mengingat bahwa penelitian ini akan lebih berat pada preprocessingnya. Contoh: perlu tidaknya skew correction untuk menangani penulisan aksara miring (yang diproses tulisan tangan). Perjelas juga bagaimana cara segmentasi dilakukan. Apalagi tulisan yang diproses tulisan anak SD. Kecuali jika tulisan yang diproses adalah tulisan cetak.
2. Pertimbangkan untuk menggunakan klasifikasi yang sudah ada. Fokus digeser pada preprocessing saja (saran).

Tanggapan :

1. Sesuai petunjuk dari penguji, pada tahap preprocessing ditambah *skew correction*. Jadi pada tahap preprocessing ini dilakukan beberapa tahap yaitu dengan cara meningkatkan kecerahan, kontras, menghilangkan nois, konversi ke citra biner, *resize*, *thinning*, dan *skew correction*. (penjelasan terdapat pada bagian preprocessing). Tulisan yang digunakan adalah tulisan siswa SMKN 1 Sumenep (penjelasan terdapat pada bagian dataset).

2. Sudah dipertimbangkan menggunakan klasifikasi yang ada sesuai review paper tapi objek penelitiannya menggunakan carakan madura.

Reviewer 5 : Judi Prajetno

Sugiono, Ir., M.M. Status :

DIPERBAIKI

Pesan : 1) pengumpulan data tidak perlu disebutkan menggunakan alat tertentu (smartphone) yang penting data yang diambil memadai sebagai syarat penelitian
2) pada tabel fitur tertulis ekstraksi ukuran "ada" , berbentuk shape energy kah atau berbentuk lain.

Tanggapan :

1. Pada bagian latar belakang dan input output sudah diperbaiki menjadi tidak menyebutkan penggunaan alat tertentu
2. Berbentuk lain

Hasil Rapat

Status : **DIPERBAIKI**

Syarat : 1. Output berupa kata
2. Input 500 data set yang ditulis oleh siswa SMK
3. Data set training 300, testing 200.
4. Fitur ada 2: panjang dan lengkungan
5. Hipotesis terlalu tinggi, berdasarkan rata-rata paper.
Silahkan diskusikan dengan pembimbing.

Tanggapan :

1. Output penelitian berupa kata huruf latin. Penjelasan terdapat pada bagian *input* dan *output* sistem. (hal 16)
2. Input berupa 500 kata carakan madura yang ditulis oleh 20 siswa SMK. Penjelasan terdapat pada bagian dataset. (hal 16)
3. Pada penelitian ini menggunakan 300 kata carakan madura sebagai data training dan 200 kata carakan madura sebagai data testing. Penjelasan terdapat pada bagian kemungkinan uji coba. (hal 16)
4. Ekstraksi fitur awalnya 2 ditambah 1 ekstraksi fitur lagi jadi ada 3 yaitu: panjang dan kelengkungan ditambah pixel population. Penjelasan terdapat pada bagian ekstraksi fitur. (hal 13)
5. Hipotesis penelitian sudah didiskusikan dengan pembimbing dan diperbaiki menjadi 70%. Penjelasan terdapat pada bagian hipotesis (hal 2)

Saran :

Tanggapan :

KONVERSI CARAKAN MADURA MENJADI HURUF LATIN MENGUNAKAN ARTIFICIAL NEURAL NETWORK DAN HIDDEN MARKOV MODEL

Nama : Miftah Farid
NRP : 215210499
Jurusan : S2 Teknologi Informasi
Pembimbing : Joan Santoso, S.Kom, M.Kom
Co. Pembimbing : DR. Ir. Endang Setyati, MT

I. Latar Belakang

Aksara berasal dari bahasa Sansekerta dari kata “a” yang berarti tidak dan “khasara” yang berarti termusnahkan. Jadi aksara mempunyai arti sesuatu yang kekal , langgeng atau tidak termusnahkan. Dikatakan “kekal” karena aksara memiliki peran mendokumentasikan dan mengabadikan suatu peristiwa ke dalam bentuk tulisan. Dari arti tersebut aksara dapat didefinisikan sebuah sistem simbol visual yang tertoreh pada suatu media, serta memiliki fungsi untuk mengungkapkan unsur-unsur yang mengekspresikan suatu bahasa. Pada aksara ada unsur-unsur lebih kecil antara lain grafem, huruf, diakritik, anda baca, dan lain-lain.

Carakan Madura merupakan bagian dari aksara Nusantara yang seringkali dikaitkan dengan inkulturisasi kebudayaan India sebelum berkembangnya agama Islam di Nusantara dan kolonialisasi bangsa-bangsa Eropa di Nusantara. Masyarakat Madura mengenal carakan sebagai lambang bunyi dan sarana penulisan itu berasal dari Jawa. Mereka menyebutnya aksara Jhaba atau Jhaban, disamping penyebutan carakan Madura.

Jumlah carakan Madura sama dengan di Jawa yaitu sebanyak 20 buah dengan urutan yang sama. Bentuk hurufnya pun sama, hanya lafalnya yang berbeda karena ada perbedaan fonologi sehingga apabila carakan madura dan aksara jawa dikonversi ke huruf latin maka akan menghasilkan huruf latin yang berbeda.

Pelajaran Bahasa Daerah merupakan pelajaran yang sebagian pembelajarannya berisi tentang carakan Madura. Untuk itu perlu dilakukan proses *handwriting recognition* untuk mempermudah kegiatan belajar mengajar tentang carakan Madura berupa konversi carakan madura ke huruf latin Indonesia. Proses tersebut diperlukan pengambilan foto sebagai arsip dalam bentuk citra digital. Namun citra digital tersebut harus dipisah antara aksara carakan Madura dan *background* sehingga hanya didapatkan carakan madura saja. Semua proses tersebut membutuhkan pengolahan citra digital yang bisa dijalankan berupa *handwriting recognition*. Kemudian aksara carakan Madura tersebut dikonversi menjadi huruf latin Indonesia sehingga mempermudah orang membacanya.

II. Tujuan dan Manfaat

Tujuan dan manfaat dari penelitian ini adalah mengkonversi carakan madura menjadi huruf latin melalui proses pengolahan citra digital dengan harapan dapat membantu proses pembelajaran siswa di sekolah.

III. Hipotesis

Hipotesis yang akan diuji dalam penelitian ini adalah metode pengenalan menggunakan *artificial neural network* (ANN) dan *Hidden Markov Model* (HMM) dengan akurasi minimal 70%.

IV. Teori Dasar Carakan Madura

Carakan Madura Terdiri dari aksara ghajang, pangangguy sowara, pangangguy panyeghek, pangangguy panamba, angghuyanna, dan lain-lain.

| | | | | |
|----|-----|----|-----|-----|
| | | | | |
| a | na | ca | ra | ka |
| | | | | |
| da | ta | sa | wa | la |
| | | | | |
| pa | dha | ja | ya | nya |
| | | | | |
| ma | ga | ba | tha | nga |

Gambar 1. Aksara ghajang, pasangan dhampeng dan ghantongan

Contoh penggunaan aksara ghajang:

= acaca = berbicara

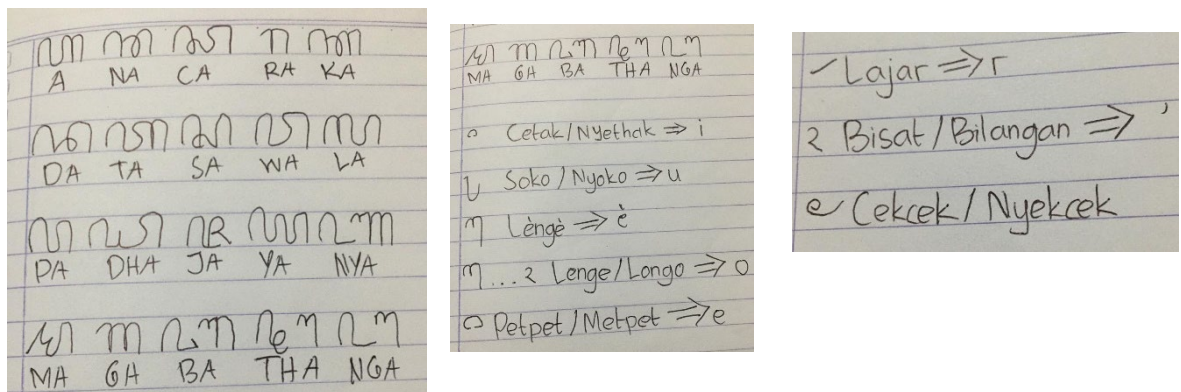
| | | | | |
|-------|-------------|-------|------|--------|
| | | | | |
| Léngé | Léngé/Longo | Cétak | Soko | Petpet |

Gambar 2. Pangangguy sowara

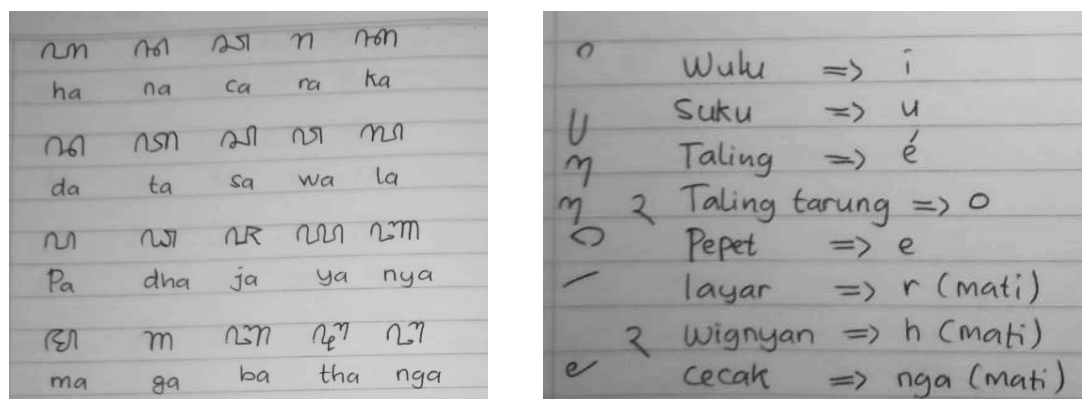
Contoh penggunaan pangangguy sowara

= buku = buku

Secara visual carakan madura dan aksara jawa tidak ada perbedaan. Berikut contoh tulisan tangan carakan madura dan aksara jawa.



Gambar 3. Contoh tulisan tangan carakan madura



Gambar 4. Contoh tulisan tangan aksara jawa

V. Tinjauan Pustaka

Berikut adalah beberapa penelitian yang melakukan pengenalan karakter dan konversi huruf:

- Judul : Information Extraction and Text Mining of Ancient Vattezhuthu Characters In Historical Documents Using Image Zoning

Penulis : E.K. Vellingiriraj, Dr. M. Balamurugun,

Penerbit : Department of Computer Science and Engineering Kongu Engineering College, Perundurai, Tamil Nadu, India
Department of Computer Science and Engineering Christ University, Bangalore, Karnataka, India

Tahun : 2016

Tujuan paper ini untuk mengembangkan sebuah sistem pengenalan karakter-karakter Brahmi, Grantha dan Vattezhuthu dari manuskrip palem dari dokumen Tamil kuno, menganalisis teks dan diterjemahkan dalam format teks digital Tamil saat ini.

Karakter Brahmi dipahat di batu, pot tanah liat, piring tembaga dll. Arsitektur keseluruhan sistem pengenalan karakter Tamil Brahmi untuk prasasti batu. Komponen utama dari sistem ini adalah:

(a) Pengambilan gambar

Gambar Prasasti Brahmi dikumpulkan. Gambar diambil dengan kamera High Definition / Digital Single Lens Reflex (HD / DSLR) berkualitas tinggi atau resolusi tinggi dan disimpan dalam format JPEG.

(b) Preprocessing

➤ Image cropping

Gambar diubah menjadi gambar abu-abu atau hitam dan putih serta menghapus noise.

➤ Segmentasi

Segmentasi dibagi menjadi tiga kelompok: segmentasi baris, segmentasi kata dan segmentasi karakter.

➤ Resize gambar

Setiap karakter harus diubah menjadi ukuran yang sama. Gambar karakter diubah ukurannya menjadi gambar 100X100 piksel.

➤ Penipisan gambar

Karakter yang tebal mengekstrak karakter yang tipis dengan menggunakan piksel gelap terdekat untuk meringankan perubahan warna ke rentang tertentu.

➤ Image binarization

Sebuah karakter menyimpan matriks Boolean yang digunakan untuk menyimpan 0 atau 1. Pixel gelap disimpan sebagai piksel 1 dan piksel terang disimpan sebagai 0 dengan teknik zonasi gambar.

(c) Data set training

Kelas karakter Brahmi yang berbeda dan Vattezhuthu dari berbagai penulis dari semua 237 karakter Vattezhuthu dikumpulkan dan disimpan dalam kumpulan data pelatihan.

(d) Pengenalan karakter

Kumpulan data karakter Vattezhuthu cocok dengan data pengguna saat ini. Algoritma deteksi tepi digunakan untuk mengidentifikasi setiap karakter dalam dataset pelatihan menggunakan zonasi gambar.

(e) Unicode text

Teks Tamil saat ini dikonversi menggunakan Unicode

(f) Pengambilan dari database

Setelah mengubah teks Unicode, ejaan dan maknanya dicek di database kamus. Jika sebuah kata tidak memberi arti dalam kalimat tersebut, artinya yang paling dekat adalah dicari dari database.

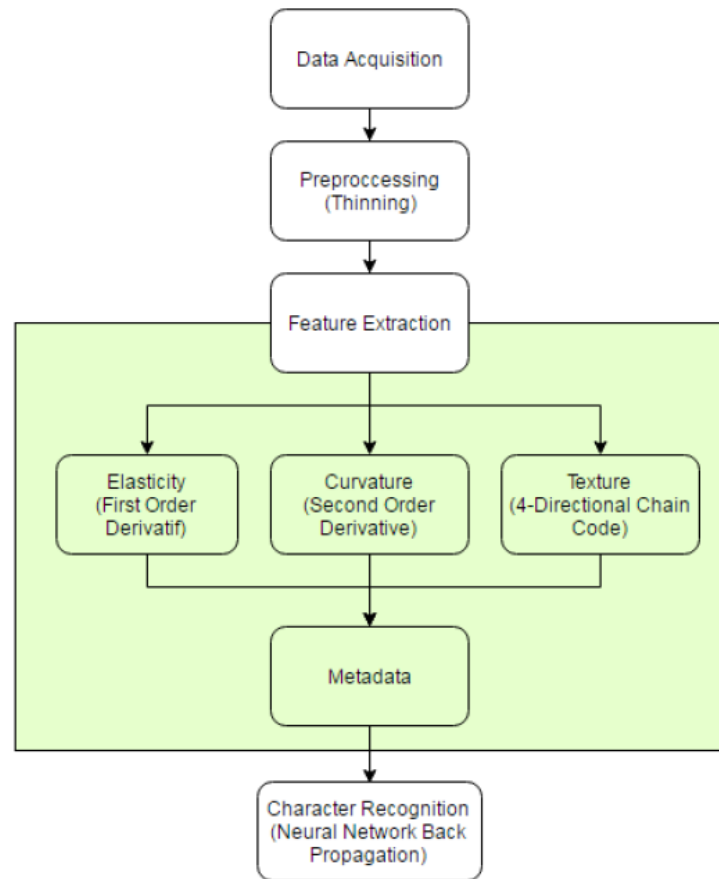
Untuk mengukur keefektifan sistem, database Vattezhuthu digunakan sebagai sumber untuk proses pelatihan dan pengujian. Database mencakup 5000 karakter yang tersimpan untuk pelatihan dan 4.000 karakter Vattezhuthu tulisan tangan yang diproduksi oleh 60 penulis, yang menulis 5 sampel dari setiap 30 kelas untuk pengujian. Karakter Vattezhuthu sudah tua, oleh karena itu penulis tidak menulis karakter uji secara jelas dan benar. Tingkat konversi vokal dalam karakter Brahmi

adalah 93,45% dan Vattezhuthu adalah 91,11%. Konsonan adalah 92,75% karakter Brahmi dan 90,75% Vattezhuthu. Vokal konsonan dari Brahmi adalah 90,24% dan Vattezhuthu adalah 89,12. Konversi karakter secara keseluruhan adalah 91,57% karakter Brahmi dan 89,75% Vattezhuthu. Jadi perbandingan hasil ini adalah karakter Brahmi set yang lebih akurat dengan Vattezhuthu.

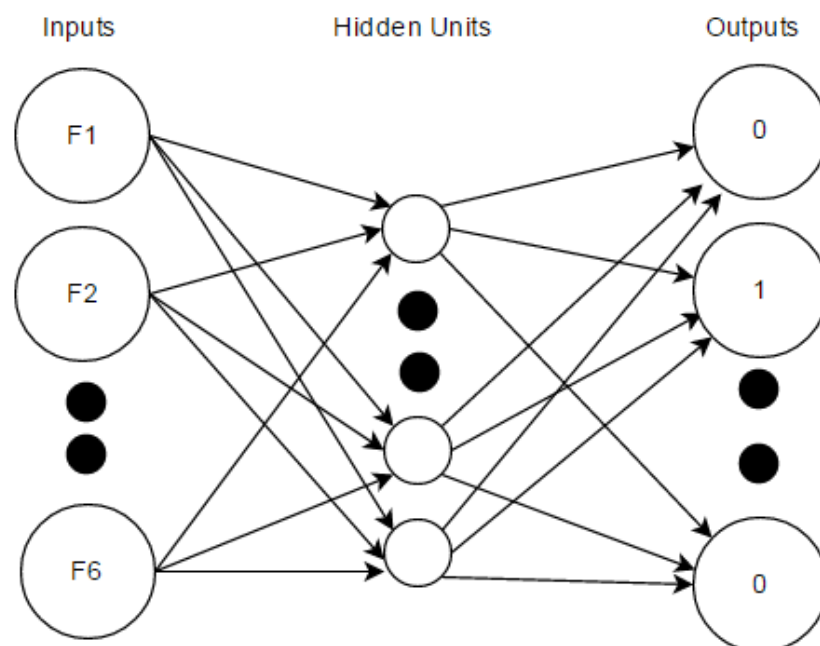
2. Judul : Feature Extraction of Character Image Using Shape Energy
Penulis : Galih Hendra Wibowo, Riyanto Sigit, Aliridho Barakbah
Penerbit : Department of Information and Computer of Engineering Graduate
Program of engineering Technology Politeknik Elektronika Negeri
Surabaya, Indonesia
Tahun : 2016

Penelitian ini mencoba untuk menciptakan alternatif baru dalam proses ekstraksi menghitung nilai energi karakter. Desain sistem dalam penelitian ini diilustrasikan pada Gambar 1. Poin penting dalam penelitian ini terletak pada ekstraksi fitur sebagai usulan metode baru..

- (a) Akuisisi data: Objeknya berupa karakter numerik atau digit dari font komputer yang didapat dari Chars74K. Dataset ini memiliki 62 kelas dengan rincian 10 kelas nomor, 26 kelas huruf kecil dan 26 kelas huruf besar.
- (b) Preprocessing: metode yang dikenal dengan Zhang-Suen digunakan dalam penelitian ini karena metode ini efisien dan cepat dalam mendapatkan kerangka karakter. dataset gambar yang digunakan memiliki kualitas gambar yang baik, sudah diformat dalam warna hitam dan putih (biner), tidak ada nois, memiliki ukuran yang sama, dan jernih.
- (c) Ekstraksi fitur: Tahap utama dalam penelitian ini adalah ekstraksi fitur menggunakan Shape Energy. Ide dasar dari metode ini adalah bagaimana membedakan antara objek yang berbeda, seperti segitiga, bujur sangkar, dan lingkaran. Perbedaan bentuk ini dapat dikategorikan sebagai tekstur objek. Selain itu, kelengkungan masing-masing objek dapat menentukan perbedaannya. Objek-objek ini memiliki kelengkungan yang berbeda, misalnya segitiga ada tiga sudut yang nilainya kurang dari 90 derajat, kuadrat terbentuk pada empat sudut dengan nilai 90 derajat, dan lingkaran yang dibentuk pada sudut-sudut yang sangat halus. Oleh karena itu, derajat kelengkungan adalah cara untuk membedakan objek yang sederhana. Dan yang terakhir adalah panjang objek itu sendiri. Dengan menggunakan skala yang sama, objek memiliki kontur panjang yang berbeda. Oleh karena itu, metode ini mencakup tiga komponen struktural, yaitu nilai panjang suatu objek dalam gambar (elastisitas), nilai kelengkungan, dan nilai tekstur.
- (d) Character recognition: Dalam penelitian ini, pengenalan karakter dilakukan dengan menggunakan neural network back-propagation



Gambar 5. Desain sistem



Gambar 6. Struktur *Neural Network*

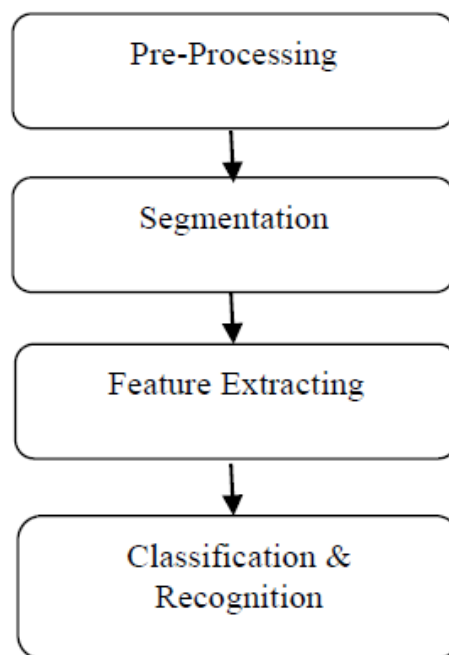
Metode baru yang diusulkan Shape Energy, dengan komponen utamanya, yaitu elastisitas, kelengkungan, dan tekstur, bisa dijadikan alternatif ekstraksi fitur karakter, terutama pada gambar. Tingkat kinerja metode ini mampu mencapai 92,25% dengan menggunakan klasifikasi jaringan syaraf tiruan.

3. Judul : Intelligent Tool For Malayalam Cursive Handwritten Character Recognition Using Artificial Neural Network And Hidden Markov Model
Penulis : Thulasi Kishna N.P, Seenia Francis
Penerbit : Department Of Computer Science and Engineering Master Of Technology Jyothi Engineering College Thrissur, Kerala

Paper ini menjelaskan hand character recognition dengan menggunakan artificial neural network (ANN) dan hidden markov model (HMM) melalui tahapan-tahapan berikut:

- (a) akuisisi gambar
Karakter tulisan tangan yang akan dikenali diambil dengan menggunakan pemindai optik
- (b) Preprocessing
Tahap preprocessing memuat koreksi kemiringan dan normalisasi. Deteksi dan koreksi kemiringan adalah langkah persiapan awal yang menentukan dalam prosedur pengenalan karakter dengan menggunakan algoritma penjarangan bersama dengan transformasi Hough kemudian estimasi kemiringan diubah dengan menggunakan metode transformasi koordinat.
- (c) Segmentasi
Di sini daerah segmentasi dikenali dari proyeksi vertikal.
- (d) Ekstraksi fitur
Fase ekstraksi fitur menggunakan ekstraksi fitur tekstur dari karakter tulisan tangan. Untuk tujuan ini, digunakan filter median.
- (e) Klasifikasi dan pengenalan
Artificial Neural Network (ANN) berfungsi untuk mengkorelasikan fitur input dan mencocokkan kelas terbaik. Hidden Markov Model (HMM) digunakan untuk mengenali atribut tersegmentasi. Setelah ekstraksi fitur, fitur yang diekstraksi dikenali menggunakan pelatihan HMM dan proses pemeringkatan yang berisi pelatihan dan gambar uji sebagai input.

Sistem ini diimplementasikan menggunakan MatLab 2014a dengan akurasi 93,4%. HMM hibrida dengan ANN memiliki peran penting dalam kinerja sistem.



Gambar 7. *System of recognition*

VI. Ruang Lingkup

A. Perancangan Sistem

Penelitian tentang pengenalan dan konversi aksara jawa pernah dilakukan oleh Heny Kunsrosita dengan judul konversi aksara jawa ke huruf latin menggunakan metode bayesian network. Penelitian tentang pengenalan dan konversi aksara jawa juga pernah dilakukan oleh Edo Prasetyo Nur Adi Wijaya dengan judul pengenalan aksara jawa cetak menggunakan kombinasi dinamic bayes dan HMM untuk itu penulis saat ini melakukan penelitian dengan menggunakan metode yang berbeda yaitu menggunakan *artificial neural network* dan *hidden markov model* dengan harapan hasil yang lebih bagus.

Tabel 1. Tabel perbandingan penelitian yang pernah dilakukan oleh Hantoro Soegiarto dan Heny Kunsrosita dengan yang akan dilakukan penulis saat ini.

| Tahapan | Heny Kunsorita | Edo Prasetyo Nur Adi Wijaya | Miftah Farid (Penulis) |
|---------|---|--|--|
| Input | Tulisan tangan 20 aksara jawa pokok dan sandangan | Font 20 aksara jawa pokok, sandhangan dan pasangan | Tulisan tangan kata carakan madura berupa 20 aksara ghajang, pasangan dhampeng dan ghantongan, pangangguy sowara, pangangguy panyeghek, pangangguy panamba |

| Tahapan | Henry Kunsorita | Edo Prasetyo Nur Adi Wijaya | Miftah Farid (Penulis) |
|-----------------|---------------------------------|-----------------------------|---|
| Preprocessing | Grayscale, filtering, threshold | Noise removal | meningkatkan kecerahan, kontras, menghilangkan noise, konversi ke citra biner, <i>resize</i> , <i>thinning</i> , dan <i>skew correction</i> |
| Segmentasi | Horizontal integral projection | Histogram dan CCL | <i>Vertical, connected component labeling</i> |
| Ekstraksi Fitur | Thinning, chain code generator | Dinamik Bayes dan HMM | Elastisitas dan lengkungan |
| Klasifikasi | Bayesian network | - | Hibrid <i>artificial neural network</i> dan <i>hidden markov model</i> |
| Output | Suku kata huruf latin | Kata huruf latin | Kata huruf latin |

Karakter yang dikenali pada penelitian ini yaitu tulisan tangan aksara ghajang, pasangan dhampeng dan ghangtongan, pangangguy sowara, pangangguy panyeghek, pangangguy panamba.

| | | | | |
|----|-----|----|-----|-----|
| | | | | |
| a | na | ca | ra | ka |
| | | | | |
| da | ta | sa | wa | la |
| | | | | |
| pa | dha | ja | ya | nya |
| | | | | |
| ma | ga | ba | tha | nga |

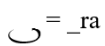
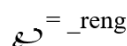
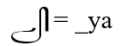

Gambar 8. Aksara ghajang, pasangan dhampeng dan ghangtongan

| | | | | |
|-------|-------------|-------|------|--------|
| | | | | |
| Léngé | Léngé/Longo | Cétak | Soko | Petpet |

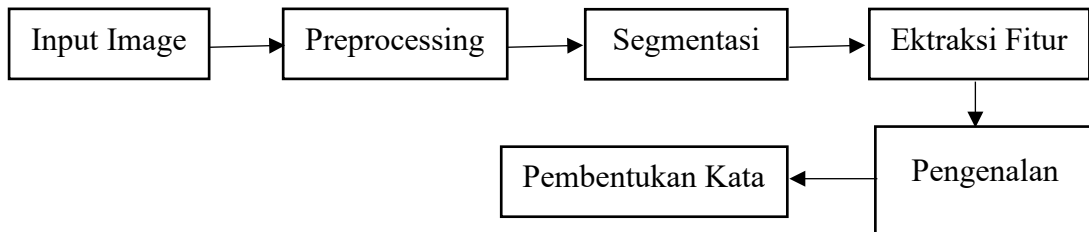
Gambar 9. Pangangguy sowara

| | | |
|----------------|-------|----------------|
| | | |
| Bisat/Bilangan | Lajar | Cekcek/Nyekcek |

Gambar 10. Pangangguy panyeghek

| | | | |
|---|--|---|--|
|  Cakra/Pedher |  Kerret/Perper |  Soko Malja/ Pengkal |  Papaten/Pangkon |
|---|--|---|--|

Gambar 11. Panganggu panamba

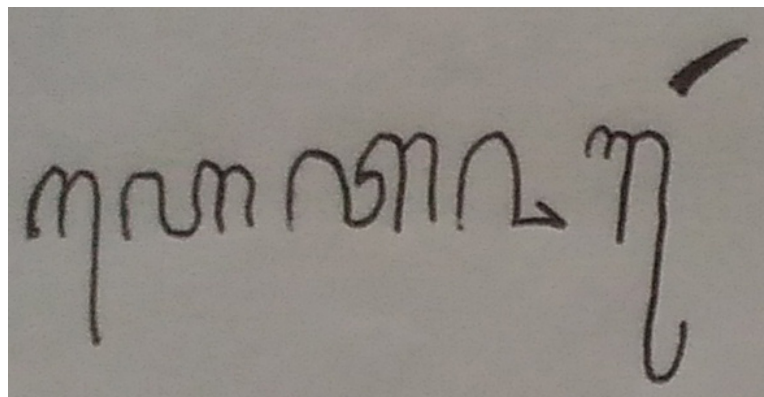


Gambar 12. Blok diagram desain sistem

Gambar 4 menunjukkan tahapan-tahapan yang akan dilakukan pada penelitian ini. Berikut ini adalah penjelasan dari masing-masing tahapan pada blok sistem:

1. Input Image

Pengambilan citra dilakukan dengan menggunakan kamera. gambar tersebut akan terbentuk dalam format JPG dengan tulisan berwarna hitam dan background berwarna putih.

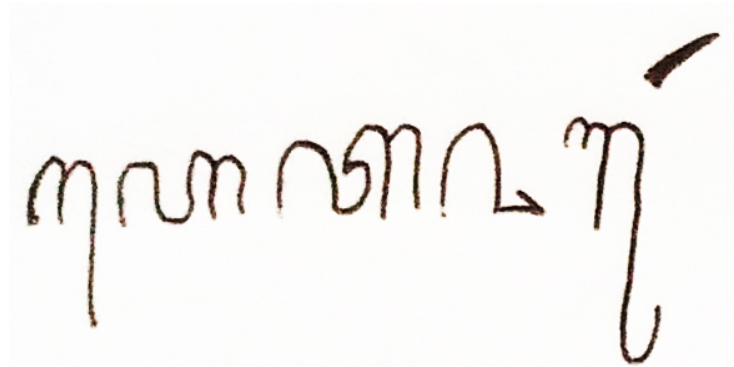


Gambar 13. Input image

2. Preprocessing

Pada tahap preprocessing ini dilakukan beberapa tahap yaitu dengan cara meningkatkan kecerahan, kontras, menghilangkan noise, konversi ke citra biner, *resize*, *thinning*, dan *skew correction*. Peningkatan kecerahan dilakukan dengan cara menambahkan suatu konstanta terhadap seluruh nilai pixel. Kontras dalam suatu citra menyatakan distribusi warna terang dan gelap. Suatu citra berskala keabuan dikatakan memiliki kontras rendah apabila distribusi warna cenderung pada jangkauan aras keabuan yang sempit. Sebaliknya, citra mempunyai kontras tinggi apabila jangkauan aras keabuan lebih terdistribusi secara lebar. Kontras dapat diukur berdasarkan perbedaan antara nilai intensitas tinggi dan nilai intensitas rendah yang menyusun pixel-pixel dalam citra. Kemudian menghilangkan noise pada citra. Citra berskala keabuan perlu dikonversi ke citra biner sehingga citra hanya berwarna hitam dan putih.

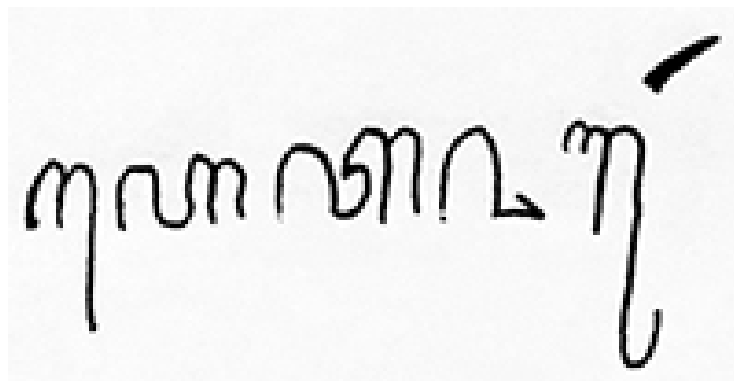
Resize gambar perlu dilakukan untuk menyamakan ukuran citra dengan cara digitalisasi citra biner dan mempercepat proses pengolahan citra. Penipisan gambar dilakukan dengan cara *thinning*^[7]. Kemudian dilakukan *skew correction* untuk memperbaiki kemiringan tulisan.



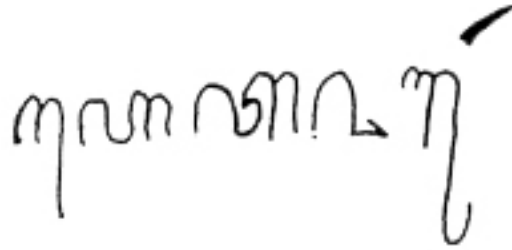
Gambar 14. Contoh peningkatan kecerahan, kontras dan nois



Gambar 15. Contoh konversi ke citra biner



Gambar 16. Contoh *resize*



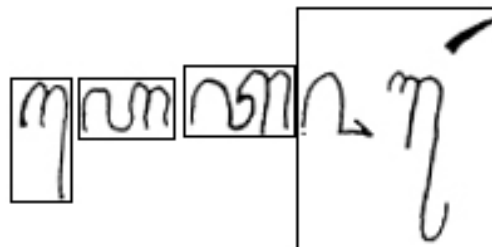
Gambar 17. Contoh *thinning*



Gambar 18. Contoh *skew correction*

3. Segmentasi

Segmentasi merupakan proses yang dilakukan untuk memisahkan komponen setiap teks. Segmentasi yang digunakan yaitu segmentasi karakter (vertikal)^[1]. Kemudian dilanjutkan dengan proses pelabelan aksara. Proses ini digunakan untuk memberikan label yang berbeda pada setiap karakter sehingga karakter yang satu dengan karakter yang lain dapat dipisahkan berdasarkan label yang dimilikinya. Proses pelabelan aksara menggunakan *connected component labeling*. *Connected component labeling* dilakukan dengan memeriksa suatu citra dan mengelompokkan setiap *pixel* kedalam suatu komponen terhubung menurut aturan keterhubungan (4, 8, atau *m-connectivity*). Setiap komponen terhubung yang saling tidak terhubung (*disjoin*) pada suatu citra akan diberi label yang berbeda.



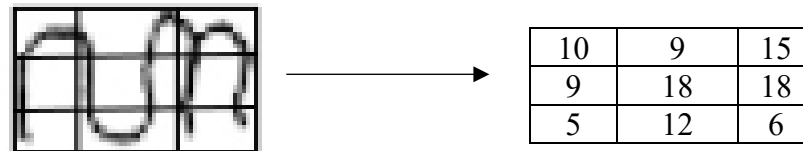
Gambar 19. Contoh segmentasi karakter

4. Ekstraksi Fitur

Proses ini digunakan untuk memilih informasi dari ciri yang ada dengan menggunakan metode:

a. Pixel population ratio

Setiap citra karakter dibagi menjadi 3x3 bagian sehingga menjadi 9 bagian. Setiap bagian dalam keseluruhan citra akan dicari piksel banyaknya objek.



Gambar 20. Contoh pixel population

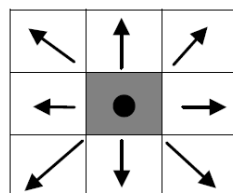
b. Rasio panjang banding tinggi

Pada fitur ini menggunakan rasio perbandingan antara panjang dengan tinggi huruf. *Length* merupakan jarak antara ujung dan pangkal huruf. *Width* merupakan jarak yang tegak lurus pada huruf

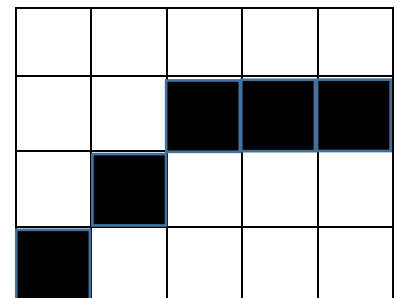
$$\frac{\text{length}}{\text{width}}$$

c. Fitur lengkungan

Fitur lengkungan menggunakan metode chain code hanya pada bagian yang melengkung saja untuk mengenali bentuk karakter yang berbeda dengan cara mendeteksi dan mengambil arah pergerakan citra sesuai dengan arah chaincode



| | | |
|---|---|---|
| 3 | 2 | 1 |
| 4 | | 0 |
| 5 | 6 | 7 |



1, 1, 0, 0, 0

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| 3 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |

Gambar 21. Contoh ekstraksi fitur

5. Pengenalan Karakter

Setelah ekstraksi fitur, fitur yang diekstraksi dikenali menggunakan *Artificial Neural Network (ANN)*. *ANN* akan mengkorelasikannya dengan fitur-fitur *input* serta mengelompokkannya dalam kelas yang cocok.

| No. Urut | Gambar Karakter | Matriks Ciri | | | ANN |
|----------|-----------------|--------------|----|----|-----|
| 1 | | 6 | 9 | 6 | |
| | | 0 | 0 | 5 | |
| | | 0 | 0 | 5 | |
| 2 | | 10 | 9 | 15 | |
| | | 9 | 18 | 18 | |
| | | 5 | 12 | 6 | |
| 3 | | 10 | 9 | 15 | |
| | | 9 | 22 | 18 | |
| | | 5 | 12 | 6 | |
| 4 | | 0 | 0 | 6 | |
| | | 20 | 15 | 10 | |
| | | 0 | 0 | 11 | |

Tabel 2. Contoh *output* pengenalan karakter

6. Pembentukan Kata

Setelah pengenalan, outputnya dikonversi dengan unicode. Kemudian untuk pembentukan kata menggunakan *Hidden Markov Model (HMM)*.

→ lenge a ta bu layar → etabur

Gambar 22. Contoh hasil *output*

B. Batasan Penelitian

Batasan pada penelitian tesis ini adalah sebagai berikut:

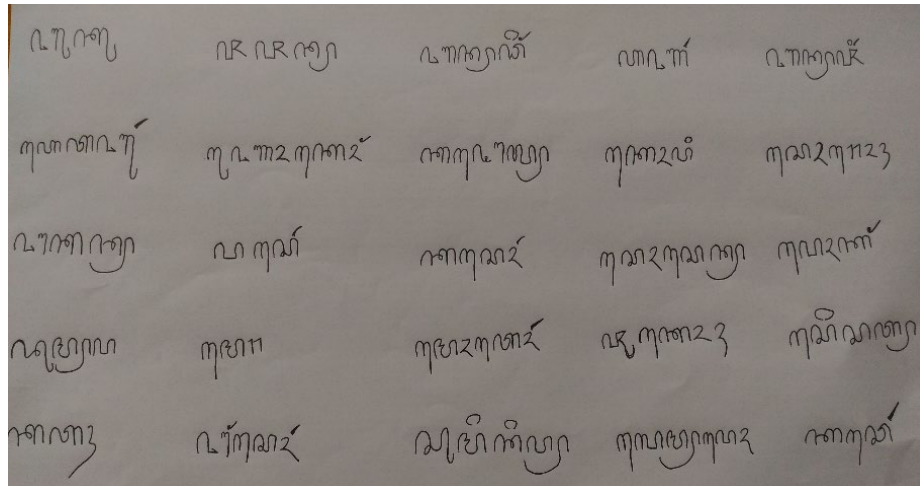
1. Konversi hanya pada carakan Madura
2. Outputnya berupa kata huruf latin
3. Input berupa format .JPG

Tabel 3. Perbandingan penelitian yang sebelumnya dengan penelitian yang sekarang

| No | Fitur | Penelitian Sebelumnya | Penelitian Sekarang |
|----|---------------------------|--|--|
| 1 | Jenis huruf yang diteliti | - karakter-karakter Brahmi, Grantha dan Vattezuthu - Huruf dan angka latin - Malayalam | Tulisan tangan kata carakan madura berupa 20 aksara ghajang, pangangguy sowara, pangangguy panyeghek, pasangan dhampeng dan ghantongan |
| 2 | Ekstraksi ukuran | Shape energy | ada |
| 3 | Ekstraksi warna | Tidak ada | Tidak ada |
| 4 | Ekstraksi bentuk | Shape energy | ada |
| 5 | Ekstraksi texture | Shape Energy | Tidak ada |
| | | Filter median | Tidak ada |
| 6 | Klasifikasi | Artificial Neural Network hibrida Hidden Markov model | ada |

C. Dataset

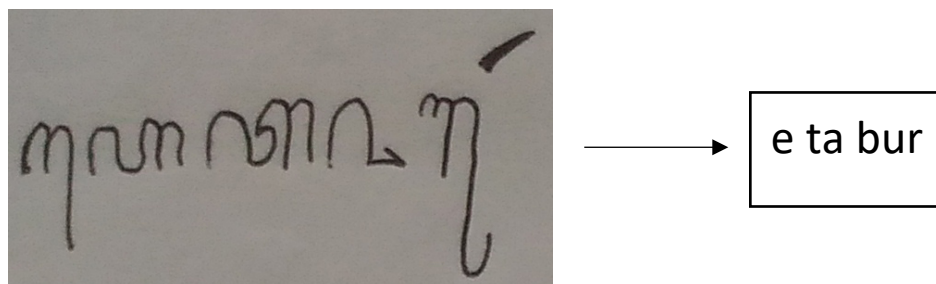
Dataset yang digunakan dalam penelitian ini adalah data yang diambil dari tulisan tangan siswa SMKN 1 Sumenep diatas kertas warna putih dan pulpen warna hitam dengan dataset 20 siswa yang dimana setiap siswa menulis 25 kata. Jadi keseluruhan data set ada 500 kata carakan madura



Gambar 23. Tulisan tangan seorang siswa

D. Input dan Output Sistem

Input dalam penelitian ini berupa tulisan tangan satu kata carakan madura berupa image dalam format .JPG. Untuk output penelitian ini adalah berupa kata huruf latin Indonesia dari input.



Gamabar 24. Input output sistem

E. Kemungkinan Uji Coba

Dari penelitian ini 300 kata berupa tulisan tangan yang akan digunakan sebagai data training dan 200 kata berupa tulisan tangan akan digunakan sebagai data testing.

VII. Metode Penelitian

Metode pada penelitian ini meliputi:

A. Studi Literature

Studi literatur dalam penelitian ini adalah mencari referensi yang relevan yang sesuai dengan permasalahan yang diangkat dalam penelitian yang diangkat dalam penelitian ini. Seperti teori tentang carakan Madura dan metode-metode yang digunakan dalam penelitian ini. Referensi dalam penelitian ini diambil dari buku, jurnal, aritkel, laporan penelitian, dan situs-situs internet.

B. Pengumpulan Data

Dataset yang digunakan dalam pelatihan dan pengujian didapat dari hasil pengambilan gambar dari tulisan carakan madura siswa-siswi SMKN 1 Sumenep.

C. Perancangan

Pada tahapan ini dilakukan proses perancangan metode, perangkat lunak dan bahasa pemrograman konversi yang akan digunakan.

D. Implementasi dan Pengujian

Pada tahapan ini dilakukan implementasi konversi huruf serta melakukan pengujian menggunakan citra uji untuk mendapatkan kesimpulan dan saran pengembangan penelitian lebih lanjut.

E. Penyusunan Laporan

Pada tahapan ini dilakukan penyusunan laporan yang menjelaskan mengenai dasar teori dan metode yang digunakan dalam penelitian yang berhasil dilakukan.

VIII. Jadwal Penelitian

Untuk menyelesaikan penelitian ini dibutuhkan waktu selama 5 bulan. Berikut ini adalah jadwal penelitian yang akan dilakukan:

| Kegiatan | Maret | | | | April | | | | Mei | | | | Juni | | | | Juli | | | |
|--------------------|-------|--|--|--|-------|--|--|--|-----|--|--|--|------|--|--|--|------|--|--|--|
| | | | | | | | | | | | | | | | | | | | | |
| Studi Literatur | | | | | | | | | | | | | | | | | | | | |
| Pengumpulan Data | | | | | | | | | | | | | | | | | | | | |
| Perancangan | | | | | | | | | | | | | | | | | | | | |
| Ujicoba | | | | | | | | | | | | | | | | | | | | |
| Penyusunan Laporan | | | | | | | | | | | | | | | | | | | | |

IX. Daftar Pustaka

1. E.K. Vellingiraj, Dr. M. Balamurugan, Dr. P. Balasubbramanie, "Information Extraction and Text Mining of Ancient Vattezhutu Characters in Historical Documents Using Image Zoning", IEEE, 978-1-5090-0922-0/16, International Conference on Asian Language Processing, 2016
2. Galih Hendra Wibowo, Riyanto Sigit, Aliridho Barakbah, "Feature Extraction of Character Image using Shape Energy", IEEE, 978-1-5090-1640-2/16, 2016
3. Thulasi Kishna N.P, Seenia Francis, "Intelligent Tool For Malayalam Cursive Handwritten Character Recognition Using Artificial Neural Network And Hidden Markov Model", IEEE Xplore Compliant-Part Number: CFP17L34-ART, ISBN: 978-1-5386-4031-9, Proceedings of the International Conference of the International Conference on Inventive Computing and Informatics (ICICI 2017), 2017
4. Trip Tinder Pal Kaur, Dr. Naresh Garg, "Optimized Gurmukhi Text Recognition from Signboard Images Captured By Mobile Camera Using Structural Features", IEEE, 978-1-4673-6994-7/15, 2015
5. Utroq Triea. 2014. Istilah 'Aksara' Berasal dari Bahasa Sansekerta yang Berarti Tidak Musnah. <http://ensiklo.com/2014/09/12/istilah-aksara-berasal-dari-bahasa-sanskerta-yang-berarti-tidak-musnah/>
Diakses Tanggal: 1 Maret 2019
6. Bastari, Yoesi Ika Fiandarti, *Kosa Kata Bahasa Madura*, Karya Simpati Mandiri, 2009.
7. Abdul Kadir, Adhi Susanto, *Teori dan aplikasi Pengolahan Citra Digital*, CV Andi Offset, 2013.
8. Fajar Astuti Hermawati, *Pengolahan Citra Digital Konsep dan Teori*, CV Andi Offset, 2013.
9. Muhkhamad Nur Kamid, Sutejo, Researchgate, 2017