SKEW CORRECTION PADA DOKUMEN TULISAN TANGAN AKSARA LAMPUNG

(Skripsi)

Oleh

IRSYAAD ABINAYA NABIL HISYAM
1717051040

FAKULTAS ILMU MATEMATIKA PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2022

SKEW CORRECTION PADA DOKUMEN TULISAN TANGAN AKSARA LAMPUNG

Oleh

IRSYAAD ABINAYA NABIL HISYAM

Skripsi

Sebagai Salah Satu Syarat untuk Memperoleh Gelar
SARJANA ILMU KOMPUTER

Pada

Jurusan Ilmu Komputer
Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung

JURUSAN ILMU KOMPUTER

FAKULTAS ILMU MATEMATIKA PENGETAHUAN ALAM

UNIVERSITAS LAMPUNG

2022

ABSTRAK

SKEW CORRECTION PADA DOKUMEN TULISAN TANGAN AKSARA LAMPUNG

Oleh

IRSYAAD ABINAYA NABIL HISYAM

Aksara Lampung menjadi ciri khas dari Provinsi Lampung bersamaan dengan Menara Siger. Dalam usaha pelestariannya, aksara Lampung bisa dijumpai pada lambang kabupaten/kota/provinsi, papan nama jalan, plat nomor rumah, dekorasi rumah, surat undangan pesta adat, sampai usaha ekonomi kreatif, misalnya jam tangan, kemasan produk lokal, dan lain-lain. Perlombaan menulis dan membaca aksara Lampung rutin digelar, baik oleh pihak pemerintah daerah maupun swasta. Semua sekolah di Provinsi Lampung diwajibkan mengajarkan muatan lokal Bahasa dan Aksara Lampung. Pelestarian Aksara Lampung melalui teknologi informasi dilakukan dengan menjadikannya sebagai objek penelitian. Dalam pengenalan karakter, terdapat pra-proses untuk membantu program membaca tulisan dan dokumen, salah satunya adalah koreksi kemiringan (*skew correction*) yang digunakan untuk meluruskan tulisan tangan yang miring. Dengan dataset tulisan tangan aksara lampung, digunakan *library* OpenCV pada Python untuk membaca dokumen dan metode Hough Transform untuk mendeteksi sudut dan menentukan kemiringan pada tulisan dan dokumen. Setelah mendapat sudut kemiringan, program memutar dokumen sesuai dengan kemiringan yang terdeteksi. Jika kemiringan akhir berada di bawah ambang batas sudut kemiringan (threshold) 5°, maka disimpulkan bahwa program dapat memperbaiki kemiringan pada dokumen tersebut. Akurasi program ketika digunakan pada dataset tersebut adalah 50%. Program ini memiliki kekurangan dalam membaca tulisan yang samar atau hilang, dan karakter aksara Lampung yang miring sehingga mengelabui program. Banyak metode lain yang dapat digunakan untuk deteksi dan koreksi kemiringan dokumen. Untuk mendapatkan gambar yang lebih jelas dan mudah dibaca dapat dilakukan proses penebalan tulisan.

Kata Kunci: Aksara Lampung, *Skew Correction*, Python, OpenCV, *Text-Line Extraction*, *Hough Transform*

ABSTRACT

SKEW CORRECTION IN LAMPUNG SCRIPT HANDWRITTEN DOCUMENTS

By

IRSYAAD ABINAYA NABIL HISYAM

The Lampung script is an identity of Lampung Province along with the Siger Tower. To preserve it, the Lampung script can be found on the emblems of regencies/cities/provinces, to creative economy businesses, such as watches, local product packaging, and many others. All schools in Lampung Province are required to teach Lampung Language and Script. The preservation of Lampung Script through information technology is carried out by making it an object of research. In character recognition, there are pre-processes to help the program read writing and documents, one of which is skew correction which is used to straighten oblique handwriting. With the handwritten dataset of Lampung script, the OpenCV library in Python used to read documents and the Hough Transform method to detect angles and determine the slope in documents. After getting the angle of inclination, the program rotates the document accordingly. If the final slope is below the threshold of the angle of inclination (threshold) of 5°, then it is concluded that the program can correct the slope on the document. The accuracy of the program when used on the dataset is 50%. This program has flaws in reading vague or missing writing, and the character of the Lampung script is skewed so that it deceives the program. Many other methods can be used for the detection and correction of the slope of the document. To get a clearer and easier-to-read image, the process of bolding the writing can be carried out.

Keywords: Lampung Script, Skew Correction, Python, OpenCV, Text-Line Extraction, Hough Transform

Judul Skripsi	:	*SKEW CORRECTION* PADA DOKUMEN TULISAN TANGAN AKSARA LAMPUNG
Nama Mahasiswa	:	Irsyaad Abinaya Nabil Hisyam
Nomor Pokok Mahasiswa	:	1717051040
Jurusan	:	Ilmu Komputer
Fakultas	:	Matematika dan Ilmu Pengetahuan Alam

MENYETUJUI

1. Komisi Pembimbing

Dr. rer. nat. Akmal Junaidi, M. Sc.	Rizky Prabowo, S.Kom., M.Kom.
NIP 197101291997021001	NIP 198808072019031011

2. Ketua Jurusan Ilmu Komputer

Didik Kurniawan, S.Si., M.T.

NIP 198004192005011004

MENGESAHKAN

1. Tim Penguji

Ketua	**: Dr. rer. nat. Akmal Junaidi, M. Sc.**	.....
Penguji I Sekretaris	**: Rizky Prabowo, S.Kom., M.Kom**	.....
Penguji II Penguji Pembahas	**: Aristoteles, S.Si., M.Si.**	.....

2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam

KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI UNIVERSITAS LAMPUNG FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

Dr. Eng. Suripto Dwi Yuwono, S.Si., M.T.

NIP 197407052000031001

Tanggal Lulus Ujian Skripsi: 8 Februari 2022

PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama: Irsyaad Abinaya Nabil Hisyam

NPM: 1717051040

Dengan ini menyatakan bahwa skripsi saya yang berjudul "*SKEW CORRECTION* PADA DOKUMEN TULISAN TANGAN AKSARA LAMPUNG" adalah benar hasil karya sendiri dan bukan orang lain. Seluruh tulisan yang tertuang dalam skripsi ini telah mengikuti kaidah penulisan karya ilmiah Universitas Lampung. Jika di kemudian hari terbukti skripsi saya adalah hasil penjiplakan atau dibuat oleh orang lain, maka saya bersedia menerima sanksi sesuai dengan peraturan yang berlaku.

Bandar Lampung, 20 September 2022

Penulis

Irsyaad Abinaya Nabil Hisyam

NPM. 1717051040

RIWAYAT HIDUP

Penulis dilahirkan di Jakarta pada tanggal 2 Juli 2000 sebagai anak pertama dari 4 bersaudara dari pasangan Bapak Yunianto dan Ibu Ratna Ningsih. Penulis menyelesaikan pendidikan Sekolah Dasar (SD) di SDIT Al-Ishmah pada tahun 2012. Kemudian pendidikan Sekolah Menengah Pertama (SMP) di SMPI Pb. Soedirman yang diselesaikan pada tahun 2014. Kemudian melanjutkan pendidikan Sekolah Menengah Atas (SMA) di SMAN 98 Jakarta yang diselesaikan pada tahun 2017.

Penulis terdaftar sebagai mahasiswa Jurusan Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung pada tahun 2017 melalui jalur Seleksi Bersama Masuk Perguruan Tinggi Negeri (SBMPTN). Selama menjadi mahasiswa, penulis melakukan beberapa kegiatan, antara lain:

1. Menjadi anggota AMAR Rohani Islam (Rois) FMIPA pada periode 2017/2018.
2. Menjadi anggota pengurus di Bidang Media Informasi Himpunan Mahasiswa Jurusan Ilmu Komputer (Himakom) pada periode 2017/2018.
3. Menjadi Asisten Dosen Jurusan Ilmu Komputer pada tahun 2018 sampai dengan 2021.
4. Menjadi Kepala Bidang Media Informasi Rohani Islam (Rois) FMIPA pada periode 2019.
5. Menjadi Ketua Komisi 1 Legislasi untuk Dewan Perwakilan Mahasiswa (DPM) pada periode 2020.

6. Menjadi *Tour Guide* untuk mahasiswa Universitas Malaya pada acara *Student Exchange* yang dilaksanakan oleh Fakultas Matematika dan Ilmu Pengetahuan Alam pada tahun 2019.
7. Mengikuti program *Student Exchange* yang dilaksanakan oleh Fakultas Matematika dan Ilmu Pengetahuan Alam ke Universiti Putera Malaisya (UPM) pada tahun 2019.
8. Melaksanakan Kerja Praktik di Pt. Sarana Prima Telematika, Jakarta, pada tahun 2020.
9. Melaksanakan Kuliah Kerja Nyata (KKN) Periode II tahun ajaran 2020/2021.

MOTTO

“Be strong, don’t look behind. Plan for the worst, do for the best”

(Ayah)

“Sertakan segala usaha atas nama Allah. Paksa diri untuk selalu ke arah yang lebih baik, menjadi hamba Allah yang kaaffah”

(Ibu)

PERSEMBAHAN

Alhamdulillahirobbilalamin

Puji syukur kepada Allah SWT yang telah memberikan rahmat dan hidayah-Nya sehingga dapat menyelesaikan penulisan skripsi ini. Sholawat dan salam saya sanjungkan kepada Nabi Muhammad SAW.

Saya persembahkan karya ini kepada:

Ibu dan Ayah

Skripsi ini saya persembahkan untuk ibu dan ayah yang telah mengisi dunia saya dengan begitu banyak kebahagiaan sehingga seumur hidup tidak cukup untuk menikmati semuanya. Terima kasih atas semua cinta yang telah ibu dan ayah berikan kepada saya.

Adik-adikku, Yasmin, Ammara, dan Farel

Terima kasih telah memberikan semangat, dukungan, dan doa yang tak berhenti.

Seluruh Keluarga Besar, Sahabat, dan Teman-teman yang selalu memberikan semangat dan dukungan.

Almamater Tercinta, Universitas Lampung

SANWACANA

Puji syukur kehadirat Allah SWT, karena telah memberikan rahmat dan hidayahNya kepada saya sehingga saya dapat menyelesaikan skripsi dengan judul "*Skew Correction* pada Dokumen Tulisan Tangan Aksara Lampung". Saya berharap skripsi ini dapat menambah pengetahuan pagi pembaca tentang aksara dan budaya Lampung, serta *Character Recognition* dan penerapannya dalam kehidupan sehari-hari

Berbekal iman, ketabahan, dan kemauan yang kuat, atas ridho dan hidayah dari Allah *Subhanahu Wa Ta'ala,* serta bantuan dari berbagai pihak, penulis dapat menyelesaikan penelitian ini. Penulis menyadari bahwa dalam penulisan skripsi ini masih terdapat kekurangan karena keterbatasan dan pengetahuan yang peneliti miliki. Melalui kesempatan ini, penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada semua pihak yang telah memberikan dukungan moril dan spiritual.

Selama proses penulisan skripsi ini tidak terlepas dari dukungan banyak pihak yang telah membimbing, membantu dan memberi semangat kepada saya, sehingga pada kesempatan ini saya ingin menyampaikan ungkapan terima kasih kepada:

1. Ibu, ayah, adik-adik dan keluarga besar saya yang selalu tidak berhenti mendoakan, menyemangati, memodalkan, serta mendukung saya. Terima kasih atas doa yang kalian berikan untuk keberhasilan dan kesuksesan saya.
2. Bapak Dr. rer. nat. Akmal Junaidi, M.Sc. sebagai pembimbing utama yang telah membimbing saya, memberi kritik dan saran serta membina dalam menyelesaikan skripsi ini.
3. Bapak Rizky Prabowo, S.Kom., M.Kom. sebagai pembimbing kedua yang telah membimbing saya dalam memberikan ide, kritik, saran sehingga skripsi ini dapat diselesaikan dengan baik.
4. Bapak Aristoteles, S.Si., M.Si. sebagai pembahas dan selaku pembimbing akademik yang telah memberikan masukan yang sangat bermanfaat dalam penulisan dan perbaikan skripsi ini.

5. Bapak Dr. Suripto Dwi Yuwono, M.T. selaku Dekan FMIPA Universitas Lampung.
6. Bapak Didik Kurniawan, S.Si., M.T., selaku Ketua Jurusan Ilmu Komputer Universitas Lampung. iii
7. Bapak Dr. rer. nat. Akmal Junaidi, M. Sc. selaku sekretaris Jurusan Ilmu Komputer FMIPA Universitas Lampung.
8. Bapak dan Ibu Dosen Jurusan Ilmu Komputer FMIPA Universitas Lampung yang telah memberikan ilmu dan pengalaman dalam hidup untuk menjadi lebih baik.
9. Kak Zainuddin, Ibu Ade Nora Maela, dan Kak Ardi Novalia yang telah membantu segala urusan administrasi penulis di Jurusan Ilmu Komputer.
10. Teman-teman seperjuangan baik dari *discord* “D ikick Dar I Kehidupan” maupun sahabat-sahabat Receh yang selalu mendukung dan saling membantu satu sama lain.
11. Teman-teman Ilmu Komputer Angkatan 2017 yang menjadi keluarga satu angkatan selama menjalankan masa studi di Jurusan Ilmu Komputer Universitas Lampung.

Dalam proses penyusunan skripsi ini tentunya terdapat kekurangan yang disebabkan oleh keterbatasan pengalaman dan pengetahuan saya. Saya sangat mengharapkan kritik dan saran dari para pembaca sebagai pembelajaran untuk saya. Semoga isi dari skripsi ini dapat bermanfaat bagi pihak yang membaca.

Bandar Lampung, 20 September 2022

Penulis,

Irsyaad Abinaya Nabil Hisyam
NPM. 1717051040

DAFTAR ISI

Halaman

DAFTAR ISI iv

DAFTAR TABEL vi

DAFTAR GAMBAR vii

DAFTAR *PSEUDOCODE* viii

I. PENDAHULUAN 1

1.1. Latar Belakang 1

1.2. Rumusan Masalah 2

1.3. Batasan Masalah 3

1.4. Tujuan Penelitian 3

1.5. Manfaat Penelitian 3

II. TINJAUAN PUSTAKA 4

2.1. Aksara Lampung 4

2.2. *Optical Character Recognition* (OCR) 6

2.3. S*kew Correction* dan *Baseline Detection* 6

2.4. *Hough Transform* 7

III. METODOLOGI PENELITIAN 9

3.1 Tempat dan Waktu Penelitian 9

3.2 Perangkat Pendukung 9

3.3 Sumber Penelitian 10

3.4 Alur Kerja Penelitian 10

IV. HASIL DAN PEMBAHASAN 13

4.1. Persiapan Penelitian 13

4.2. Eksperimen 19

4.3. Analisis Hasil Eksperimen 20

V. SIMPULAN DAN SARAN 26

5.1. Simpulan 26

5.2. Saran 26

DAFTAR PUSTAKA 27

LAMPIRAN 30

DAFTAR TABEL

Tabel Halaman

1. Anak Huruf Aksara Lampung (Pandey, 2016) 5
2. Data Aksara Lampung 10
3. Klasifikasi Dokumen dengan *Threshold*..... 30

DAFTAR GAMBAR

Gambar Halaman

1. Ilustrasi Tulisan Tangan Latin Tidak Mendatar..... 2
2. Dua Puluh Huruf Induk Aksara. 5
3. Contoh ilustrasi skew correction dan baseline detection: (a) gambar original, (b) gambar grayscale, (c) hasil skew correction 7
4. Ilustrasi Hough Transform 8
5. Alur Kerja Penelitian 10
6. Hasil deteksi garis lurus menggunakan Hough Transform 16
7. Plot Hough Transform pada dokumen 17
8. Ilustrasi Threshold untuk batas toleransi kemiringan 20
9. Dokumen tulisan tangan aksara Lampung yang buram 21
10. Hasil deteksi garis pada dokumen yang buram..... 21
11. Contoh dokumen yang salah deteksi kemiringan oleh program 22
12. Garis Hough dokumen 22
13. Hasil rotasi dokumen 23
14. Contoh dokumen yang memiliki obstruksi pada gambar..... 24
15. Garis Hough dokumen yang bergaris bawah 24
16. Hasil Rotasi Dokumen bergaris bawah 25

DAFTAR *PSEUDOCODE*

Pseudocode Halaman

1. Kode Program Import Library 14
2. Fungsi Rotate 15
3. Kode untuk import dan convert menjadi format png 15
4. Kode program untuk mendeteksi sudut kemiringan awal 16
5. Kode program memanggil fungsi rotate untuk memutar dokumen 17
6. Kode program untuk melihat sudut akhir dokumen yang telah dikoreksi 17
7. Kode Program show Window 18

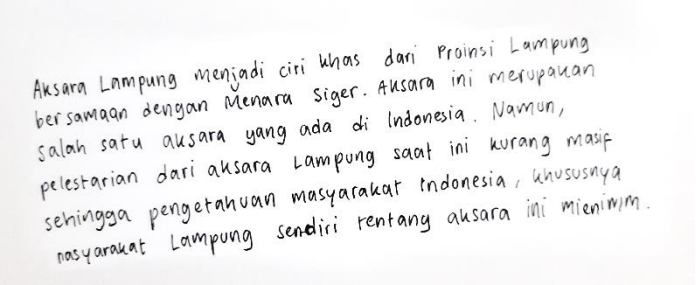
I. PENDAHULUAN

1.1. Latar Belakang

Aksara Lampung menjadi ciri khas dari Provinsi Lampung bersamaan dengan Menara Siger. Dalam usaha pelestariannya, aksara Lampung bisa dijumpai pada lambang kabupaten/kota/provinsi, papan nama jalan, plat nomor rumah, dekorasi rumah, surat undangan pesta adat, sampai usaha ekonomi kreatif, misalnya jam tangan, kemasan produk lokal, dan lain-lain. Perlombaan menulis dan membaca aksara Lampung mulai rutin digelar, baik oleh pihak pemerintah daerah maupun swasta. Semua sekolah di Provinsi Lampung diwajibkan mengajarkan muatan lokal Bahasa dan Aksara Lampung.

Pelestarian aksara Lampung melalui teknologi informasi, dilakukan dengan menjadikannya sebagai objek penelitian teknologi informasi. Berbagai kajian teknologi informasi sudah diterapkan terhadap aksara Lampung, misalnya teknik pengenalan karakter dan pemrosesan citra. *Text-line extraction* merupakan bagian penting dari pemrosesan citra yang digunakan dalam konversi teks untuk mengidentifikasi baris dari tulisan. Beberapa metode pendekatan *Text-line extraction* telah dipublikasikan, memiliki keterbatasan pada struktur halaman dokumen yang mengandung kemiringan dari baris tulisan (*text-line)*, jarak batas baris yang tidak konsisten, dan kerenggangan antar kata & permulaan paragraf yang tidak beraturan. (Kundu, 2019)**.**

Dalam *Text-Line Extraction*, langkah pertama pada tahap *preprocessing* adalah menemukan baris dokumen dan urutannya. Dengan dipisahkannya baris, maka susunan kalimat dapat diatur dengan urutan yang benar di dalam paragraf. Langkah selanjutnya adalah pengenalan aksara untuk menghasilkan transkripsi dokumen dengan urutan yang benar. Dalam dokumen tulisan tangan, tidak semua barisan tertulis dengan mendatar atau tegak lurus. Berikut adalah contoh tulisan tangan yang tidak mendatar pada Gambar 1.



Gambar 1. Ilustrasi Tulisan Tangan Latin Tidak Mendatar.

Gambar 1 memperlihatkan kemiringan (*skewness*) teks tulisan tangan yang dipindai untuk *Optical Character Recognition* (OCR). Pemindaian dokumen yang tidak disejajarkan dengan baik menghasilkan gambar digital yang miring. Salah satu solusi penyelesaian masalah ini dapat menggunakan teknik *skew correction.* Teknik ini merupakan teknik untuk memperbaiki kemiringan yang terdapat pada gambar teks tulisan tangan. Keuntungan dari teknik *skew correction* adalah tidak dibatasi pada rentang apapun dan bekerja dengan benar dengan grafik dan tabel dalam gambar teks. (Kaur, 2015)

1.2. Rumusan Masalah

Berdasarkan pemaparan latar belakang sebelumnya, maka rumusan masalah dalam penelitian ini adalah mengaplikasikan teknik *skew correction* pada dokumen tulisan tangan aksara Lampung untuk memperoleh baris tulisan yang mendatar (horizontal).

1.3. Batasan Masalah

a. Penelitian ini menggunakan aksara Lampung sebagai objek penelitian yang diperoleh dari dataset pindaian tulisan tangan aksara Lampung (Junaidi, Vajda, & Fink, 2011).
b. Penelitian ini menggunakan bahasa pemrograman Python.
c. Penelitian ini berfokus pada teknik *skew correction* dalam *preprocessing*.

1.4. Tujuan Penelitian

Penelitian ini bertujuan untuk menggunakan dan menganalisis metode *skew-correction* pada proses ekstraksi baris pada tulisan tangan aksara Lampung.

1.5. Manfaat Penelitian

Manfaat penelitian ini adalah membantu proses *preprocessing* dalam mengidentifikasi dan memperbaiki kemiringan tulisan pada dokumen tulisan tangan aksara Lampung

II. TINJAUAN PUSTAKA

2.1. Aksara Lampung

Aksara Lampung merupakan salah satu aksara tradisional Indonesia yang berkembang di selatan pulau Sumatera tepatnya di Provinsi Lampung. Dalam perkembangannya aksara Lampung digunakan untuk menulis bahasa daerah Lampung dan bahasa lainnya seperti bahasa Melayu. Aksara Lampung aktif digunakan dalam sastra maupun tulisan sehari-hari masyarakat Lampung sejak pertengahan abad ke-17 hingga pertengahan abad ke-20 sebelum fungsinya berangsur-angsur tergantikan dengan huruf Latin. Hingga saat ini, aksara Lampung masih diajarkan di Provinsi Lampung sebagai bagian dari muatan lokal, dengan penerapan yang terbatas dalam kehidupan sehari-hari. Aksara Lampung diperkirakan masuk ke daerah Sumatera bagian selatan pada zaman kerajaan Sriwijaya bersamaan dengan masuknya budaya India. Oleh karena itu, aksara ini dipengaruhi oleh Huruf Arab dan Huruf Pallawa dari India Selatan. Bentuk tulisannya berjenis suku kata, seperti dalam Huruf Arab, dengan menggunakan anak huruf sebagai keterangan vokal tulisan. Aksara Lampung telah melalui perubahan dan perkembangan bahasa yang lebih sederhana, dibandingkan dengan Aksara Lampung Kuno yang jauh lebih kompleks (Pudjiastuti, 1996).

Pada tahun 2016, aksara Lampung didaftarkan ke *Unicode* oleh A. Pandey untuk digunakan dalam format digital. (Pandey, 2016) namun belum ada kabar kelanjutannya. Usaha komputerisasi Aksara Lampung telah dilakukan oleh masyarakat termasuk oleh orang Lampung sendiri dengan membuat *software* atau *font* yang bisa diaplikasikan langsung dalam pengetikan

komputer. Aksara Lampung terdiri dari 20 huruf induk dan dibaca dari kiri ke kanan seperti huruf Latin yang divisualisasikan pada Gambar 2.

Gambar 2. Dua Puluh Huruf Induk Aksara.

Anak Huruf dan Tanda Baca pada Aksara Lampung terdiri dari tiga posisi yakni: di atas, di bawah, dan di samping seperti yang dimuat dalam Tabel 1.

Tabel 1. Anak Huruf Aksara Lampung (Pandey, 2016)

Nama	Tanda	Keterangan
Terletak di atas huruf		
Bicek		Tanda vokal e
Ulan		Tanda vokal i
Ulan		Tanda vokal é
Datasan		Tanda ganti konsonan n
Rejunjung		Tanda ganti konsonan r
Terkelubang		Tanda ganti konsonan ng
Terletak di bawah huruf		
Bitan		Tanda vokal u
Bitan		Tanda vokal o
Tekelungau		Tanda vokal au
Terletak di samping huruf		
Tekelingai		Tanda vokal ai
Keleniah		Tanda ganti konsonan h
Nengen		Tanda mati suatu konsonan

2.2. *Optical Character Recognition* (OCR)

Optical Character Recognition adalah bagian dari kecerdasan buatan untuk mengenali pola dan citra komputer, baik berupa hasil cetak(*print*) maupun tulis tangan, hasil *scan* maupun foto. OCR secara luas digunakan sebagai bentuk *input* data untuk rekam informasi, seperti paspor, faktur nota, dan mutasi rekening. OCR mengkonversi gambar huruf yang diketik, ditulis tangan, atau dicetak menjadi teks yang mampu dibaca oleh mesin secara modern. (Nicomsoft, 2012)

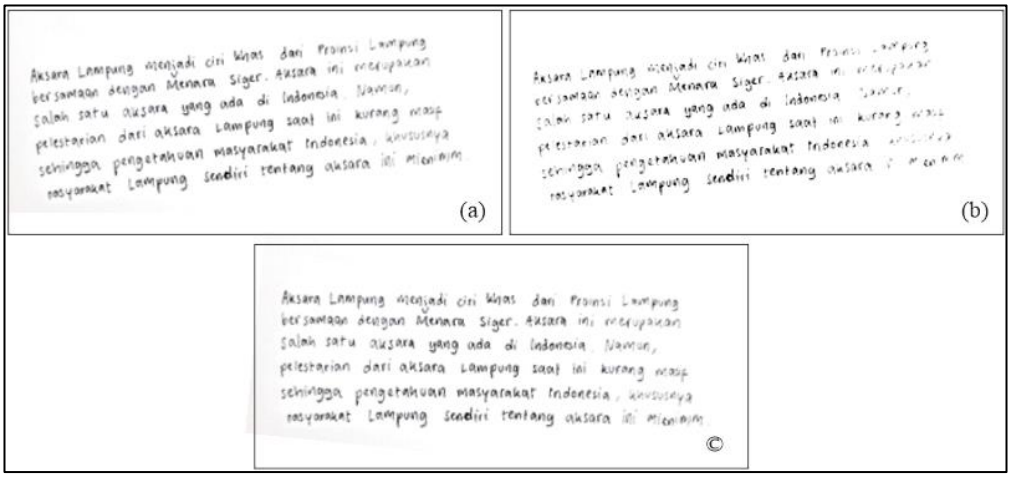
OCR model lama dilatih (*training*) dan diuji cobakan (*testing*) menggunakan gambar dari setiap karakter, dan tidak dapat bekerja jika terdapat dua *font* yang berbeda. OCR model terkini mampu menghasilkan pengenalan dengan akurasi berkisar 95-99% dan secara umum dapat membaca sebagian besar *font*. Beberapa model mampu menghasilkan *output* rapi yang hampir mendekati halaman asli, termasuk gambar, tabel, dan komponen non-tekstual lainnya. Pra-proses (*preprocessing*) biasa dilakukan pada *software* OCR untuk meningkatkan tingkat keberhasilan saat mengenali pola atau karakter. Contoh teknik pada *preprocessing* adalah *skew correction*, *slant removal*, *binarization*, dan *line extraction*.

Format gambar yang digunakan dalam penelitian ini adalah Portable Pixel Map (*.ppm*). Dokumen ppm adalah gambar berwarna 24-bit yang disusun menggunakan format teks dengan rentang piksel warna antara 0 sampai 65536. Dokumen ppm juga menyimpan data lebar dan panjang gambar, data *whitespace*, dan angka warna maksimum. Oleh karena itu, sangat mudah untuk menulis dan menganalisa program yang memproses format ini.

2.3. S*kew Correction* dan *Baseline Detection*

Skew Detection dan *Correction* adalah langkah penting dalam analisis gambar dokumen. Dalam sistem *fax*, *skew* adalah deviasi sudut dari *frame* penerima. Sudut kemiringan diperoleh dengan mencari “puncak” pada histogram dari orientasi gradien. Kemiringan (*skewness*) dari suatu dokumen atau gambar

diperbaiki dengan merotasi hingga sudut tertentu. Ada beberapa metode yang dapat digunakan dalam mendeteksi kemiringan dalam suatu halaman atau gambar yang bergantung pada komponen bersambung (*connected components*) dan menentukan sudut rata-rata yang menghubungkan *centroid*. Metode *projection histogram* bekerja dengan cara memproyeksi halaman dari beberapa sudut, dan menentukan perbedaan (*variance*) pada jumlah piksel hitam di setiap garis yang diproyeksikan. Proyeksi garis yang sejajar dengan sudut asli pada garis akan memiliki *variance* maksimum. *Baseline Detection* adalah proses mencari garis bawah dari teks pada gambar. Metode ini menghitung jumlah piksel hitam setiap baris, dan posisi dari *baseline* ditentukan dari jumlah piksel yang paling banyak (Kaur, Seema, & Kumar, 2016). Ilustrasi *skew correction* dan *baseline detection* yang dimuat pada Gambar 3.

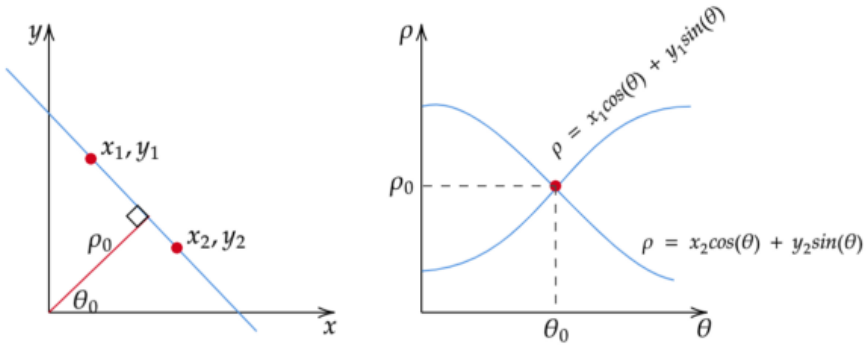


Gambar 3. Contoh ilustrasi *skew correction* dan *baseline detection*: (a) gambar original, (b) gambar *grayscale*, (c) hasil *skew correction*

2.4. *Hough Transform*

Hough Transform adalah teknik yang digunakan untuk mendeteksi garis lurus pada pemrosesan citra. *Hough Transform* bekerja dengan memetakan setiap titik pada bidang awal (z,y) ke semua titik dalam ruang (p, θ) *Hough* dari garis-garis yang melalui (2, y) dengan kemiringan 8 dan jarak p dari titik asal. Garis dominan ditemukan dari puncak di ruang Hough dan dengan demikian miring.

Salah satu batasan metode ini adalah jika teks jarang, mungkin sulit untuk memilih puncak di ruang Hough.



Gambar 4. Ilustrasi *Hough Transform*

Menggunakan gambar diatas untuk mereprentasikan Ruang Hough dengan kemiringan a dan intersep b, sekarang diwakili dengan dan di mana sumbu horizontal untuk nilai dan sumbu vertikal untuk nilai . Pemetaan titik tepi ke Ruang Hough bekerja dengan cara yang sama kecuali bahwa titik tepi (x_i, y_i) sekarang menghasilkan kurva kosinus di Ruang Hough alih-alih garis lurus. (Leavers, 1992)

Representasi normal dari sebuah garis menghilangkan masalah nilai tak terbatas dari a yang muncul ketika berhadapan dengan garis vertikal.

III. METODOLOGI PENELITIAN

3.1 Tempat dan Waktu Penelitian

3.1.1 Tempat Penelitian

Penelitian dilaksanakan di Laboratorium Komputasi Dasar Jurusan Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA) Universitas Lampung.

3.1.2 Waktu Penelitian

Penelitian dilakukan pada bulan Mei 2021 hingga penyelesaian pada bulan Oktober 2021.

3.2 Perangkat Pendukung

3.2.1 *Hardware*

Hardware yang digunakan dalam penelitian ini adalah:

a. *Processor*: *Intel Core* i5-8265U @1.60GHz up to 3.0GHz
b. *Installed* RAM: 20GB
c. Penyimpanan: 1TB

3.2.2 *Software*

Software yang digunakan pada penelitian ini adalah:

i. *Operating System*: Windows 10 Home Single 64-bit
ii. Python 3.9
iii. Library OpenCV

3.3 Sumber Penelitian

3.3.1 Data

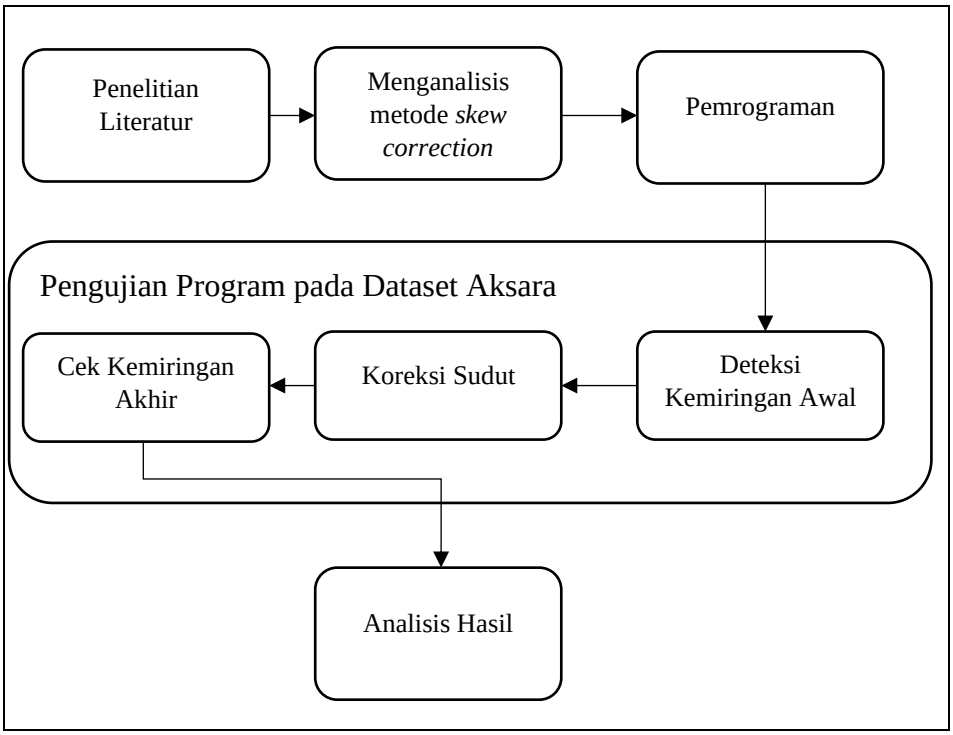
Data yang digunakan berupa gambar uji tulisan latin dan dataset aksara lampung yang terdiri dari dokumen tulisan tangan (*raw image*), komponen *grayscale* tersambung dan anotasi karakter Lampung. Dataset aksara Lampung diperoleh dari (Junaidi, Vajda, & Fink, 2011) yang dirincikan pada tabel 2

Tabel 2. Data Aksara Lampung

Jenis Data	Jumlah Data
Tulisan Latin	10 gambar
Dataset Aksara Lampung	82 gambar

3.4 Alur Kerja Penelitian

Penelitian ini melalui beberapa tahapan yang dapat dilihat pada Gambar 4.



Gambar 5. Alur Kerja Penelitian

Berdasarkan alur kerja penelitian pada Gambar 4, adapun penjelasan setiap tahap kerja sebagai berikut:

3.4.1 Penelusuran Literatur

Tahap ini adalah melakukan pencarian terhadap studi literatur untuk mendapatkan sumber data yang mendukung penelitian. Sumber data diperoleh dari berbagai sumber, baik dari penelitian sebelumnya, jurnal, *website*, atau *paper* yang mendukung permasalahan *skew correction*.

3.4.2 Menganalisis Metode *Skew Correction*

Metode *skew correction* digunakan untuk mengoreksi baris atau paragraf yang miring. Dalam penelitian ini, teknik *skew correction* yang digunakan adalah metode *projection profile* dengan batas kemiringan maksimum (*threshold)* sebesar 5°.

3.4.3 Pemrograman

Tahap ini adalah proses pembuatan kode program untuk mendeteksi kemiringan baris atau paragraf dalam tulisan dan memperbaiki nya menggunakan bahasa pemrograman *Python*. Sebelum mencoba metode diatas pada aksara lampung, terlebih dahulu akan mencoba koreksi menggunakan tulisan huruf latin.

3.4.4 Pengujian Program pada Dataset Aksara Lampung

Tahap pengujian dibagi menjadi tiga tahap, diantaranya:

3.4.4.1. Deteksi Kemiringan Awal

Tahap ini adalah mendeteksi kemiringan asal pada dokumen gambar menggunakan *library OpenCV* dan fungsi *determine_skew* pada modul *deskew*.

3.4.4.2. Koreksi Sudut

Tahap berikutnya setelah mendeteksi dan mendapatkan kemiringan awal adalah memperbaiki besar kemiringan tersebut dengan memasukan nilai kemiringan awal untuk mendapatkan hasil gambar sejajar, lalu buat dokumen baru sebagai *placeholder* untuk mendeteksi kemiringan akhir.

3.4.4.3. Cek Kemiringan Akhir

Setelah mendapat hasil gambar yang telah dikoreksi, selanjutnya adalah *periksa* kemiringan akhir dengan menggunakan fungsi *determine_skew* pada dokumen hasil akhir.

3.4.5 Analisis Hasil

Analisis hasil dilakukan untuk melihat ketercapaian program yang dibuat dapat mengoreksi kemiringan pada baris atau paragraf dalam aksara Lampung. Setelah uji coba dilakukan, selanjutnya adalah mengevaluasi program.

V. SIMPULAN DAN SARAN

5.1. Simpulan

Berdasarkan hasil penelitian, simpulan yang diperoleh pada penelitian ini adalah:

5.1.1. *Skew Correction* adalah metode untuk memperbaiki kemiringan suatu dokumen teks dalam satuan derajat yang dihitung dengan cara memutar gambar sebesar sudut kemiringan awal secara berlawanan arah. Hasilnya adalah gambar yang sejajar dimana teks menjadi tegak lurus pada dokumen, bukan miring.

5.1.2. Metode *skew correction* yang diterapkan pada barisan tulisan aksara Lampung dengan besar *threshold* 5°.

5.1.3. Hasil percobaan dapat disimpulkan bahwa akurasi yang didapatkan sebesar 50%.

5.1.4. Program ini memiliki kekurangan dalam membaca tulisan yang hilang atau tidak terlihat jelas, selain itu ada beberapa karakter aksara Lampung yang berbentuk miring sehingga program memperkirakan bahwa tulisan pada gambar tersebut tidak mendatar.

5.2. Saran

Adapun saran yang diberikan pada penelitian selanjutnya adalah sebagai berikut:

5.2.1. Dapat menggunakan metode lain untuk menangani masalah *text line extraction* atau *skew correction.*

5.2.2. Untuk memperjelas gambar yang tidak terbaca, lakukan proses penebalan tulisan.

DAFTAR PUSTAKA

DAFTAR PUSTAKA

Boukharouba, A. (2016). A New Algorithm for Skew Correction and Baseline Detection based on the Randomized Hough Transform. *Journal of King Saud University - Computer and Information Sciences*.

Fakir, M., Hassani, M., & Sodeyama, C. (2000). *ON THE RECOGNITION OF ARABIC CHARACTERS USING HOUGH TRANSFORM TECHNIQUE.*

Hull, J. (n.d.). *DOCUMENT IMAGE SKEW DETECTION: SURVEY AND ANNOTATED BIBLIOGRAPHY.*

Junaidi, A., Trianingsih, S., & Iqbal, M. (2020). Writer Identification of Lampung Handwritten documents Based on Selected Characters. *Writer Identification of Lampung Handwritten documents Based on Selected Characters*.

Junaidi, A., Vajda, S., & Fink, G. (2011). Lampung - A New Handwritten Character Benchmark: Database, Labeling, and Recognition. *Int. Workshop on Multilingual OCR.*

Kaur, R. (2015). *A REVIEW ON SKEW DETECTION AND CORRECTION IN MULTISCRIPT TEXT DOCUMENT IMAGES.* India: International Journal of Advances in Science Engineering and Technology.

Kaur, R., Seema, & Kumar, S. (2016). A Hybrid Approach To Detect And Correct A Skew In Scanned Document Images Using Fast Fourier Transform And Nearest Neighbor Algorithm.

Kundu, S. e. (2019). Text-line extraction from handwritten document images using GAN. *Expert Systems with Applications*.

Leavers, V. F. (1992). *Shape Detection in Computer Vision Using the Hough Transform.*

Nakanishi, M., & Ogura, T. (n.d.). *Real-time Line Extraction Using a Highly Parallel Hough Transform Board.*

Nicomsoft. (2012, Februari 5). Retrieved from Optical Character Recognition (OCR) – How it works: https://www.nicomsoft.com/optical-character-recognition-ocr-how-it-works/

Pandey, A. (2016). Preliminary proposal to encode the Lampung script in Unicode. 3-7.

Papandreou, A. (2011). *A Novel Skew Detection Technique Based on Vertical Projections* . Athens: International Conference on Document Analysis and Recognition.

Pudjiastuti, T. (1996). *Aksara dan Naskah Kuno Lampung dalam Pandangan Masyarakat Lampung Kini.* Jakarta: Proyek Pengkajian dan Pembinaan Nilai-nilai Budaya Direktorat Sejarah dan Nilai Tradisional Direktorat Jenderal Kebudayaan Departemen Pendidikan dan Kebudayaan.

Sun, C. (n.d.). *Skew and Slant Correction for Document Images Using Gradient Direct ion.*

Touj, S., Ben Amara, N., & Amiri, H. (2005). *Generalized Hough Transform for Arabic Printed Optical Character Recognition.*

Yan, Y., Xu, M., & Smith, J. (n.d.). *A Generalized Vertical Projection Histogram Using Multi-Plane Homology.*

Zahour, A., Taconet, B., Mercy, P., & Ramdane, S. (n.d.). *Arabic Hand-Written Text-Line Extraction.*