

Bab I

Pendahuluan

1. 1Latar Belakang

Optical Character Recognition (OCR) merupakan salah satu pengaplikasian *computer vision* yang saat ini banyak dijadikan topik penelitian karena dapat memberikan kontribusi yang cukup besar di bidang teknologi. OCR dapat mengenali dokumen (karakter, kata, dan text) dalam bentuk gambar tanpa menyalin ulang dokumen tersebut [1]. Niyazi Kilic et al [2] membuat sistem OCR untuk mengenali naskah peninggalan zaman kekaisaran Ottoman. Pada [3] Anurima Tikader et al membuat sebuah sistem pengenalan naskah di India, yaitu Inggris dan Bengali. Afroge et al [4] membuat sistem pengenalan karakter Inggris. Oleh karena itu OCR menjadi salah satu penelitian yang banyak dilakukan saat ini.

Negara Indonesia terkenal dengan kekayaan di dalamnya baik sumber daya alam maupun bahasa dan budaya. Aksara merupakan salah satu budaya Indonesia yang wajib dilestarikan agar tidak hilang keberadaannya.. Pemerintah Jawa Barat menetapkan peraturan daerah pada tahun 1996 nomor 6 yang digantikan oleh Perda No. 5 tahun 2003 mengenai pemeliharaan bahasa, sastra, dan aksara daerah.

Aksara sunda memiliki karakteristik garis yang lurus dengan lengkungan sedikit untuk mengubah arah garis. Garis – garis lurus tersebut dapat dideteksi dengan cara melihat intensitas persebaran gradien atau dengan kata lain frekuensi dari arah tepinya. Salah satu metode ekstraksi fitur pada citra adalah *histograms of oriented gradients* (HOG). Ekstraksi fitur HOG umumnya digunakan untuk mengenali objek, khususnya permasalahan deteksi pejalan kaki [5]. Metode ini mendeteksi tepi dari objek yang terdapat pada suatu citra berdasarkan gradien-gradien dari setiap piksel [4].

Support vector machine (SVM) akhir – akhir ini banyak digunakan sebagai metode klasifikasi untuk pengenalan pola, salah satunya adalah bidang OCR [1] [5]. Algoritma klasifikasi SVM berhasil tidak terjebak pada nilai minimum lokal dikarenakan ide pokok SVM adalah mencari *hyperplane* optimum dengan cara mencari jarak terbesar antara 2 kelas yang berbeda.

Pada penelitian sebelumnya mengenai pengenalan kata aksara sunda hanya sebatas pengenalan huruf aksara. Algoritma klasifikasi SVM berhasil tidak terjebak pada nilai minimum lokal karena konsep kerjanya mencari *hyperplane* paling optimum dengan cara mencari jarak terbesar antara 2 kelas yang berbeda.

Pada penelitian ini, kami mengusulkan pendekatan baru dalam mengenali Kata Aksara Sunda dengan memperhatikan karakteristik rarangkèn menggunakan metode *histograms of oriented gradients* (HOG) dan *Support vector machine* (SVM) sebagai metode klasifikasi. Penelitian

1. 2 Perumusan Masalah

Rumusan masalah pada Tugas Akhir ini adalah :

1. Bagaimana cara membangun sistem pengenalan kata aksara sunda menggunakan metode *Histograms of Oriented Gradient*?
2. Bagaimana parameter *Histograms of Oriented Gradient* yang dapat memberikan performansi paling optimal?
3. Apa jenis *font* yang dapat memberikan performansi paling tinggi?

1. 3 Tujuan

Berikut ini adalah tujuan yang ingin dicapai pada penelitian Tugas Akhir ini adalah:

1. Membangun sistem pengenalan kata aksara sunda menggunakan metode *histograms of oriented gradients*.
2. Menganalisa kinerja sistem pengenalan kata aksara sunda yang telah dibangun dengan *histograms of oriented gradients*.

1. 4 Batasan Masalah

Batasan masalah pada Tugas Akhir ini adalah data uji dan data latih berupa citra kata aksara sunda digital dengan ukuran *font* 72 dan jenis *font* yang telah disebutkan pada bab 3 mengenai *dataset*.

1. 5 Rencana Kegiatan

Rencana kegiatan yang dilakukan dalam menyelesaikan Tugas Akhir ini adalah sebagai berikut.

1. Studi literatur

Tahap pertama dalam menyelesaikan Tugas Akhir (TA) ini adalah mencari, membaca, dan memahami studi literatur yang berkaitan dengan topik pada Tugas Akhir ini. Studi literature yang dipelajari mengenai teori aksara sunda, *histograms of oriented gradient*, dan *support vector machine* jenis *multi class*. Hasil dari tahap studi literature akan digunakan sebagai acuan dalam menyelesaikan Tugas Akhir ini.

2. Pengumpulan Data

Dataset yang digunakan pada Tugas Akhir (TA) ini berupa citra digital (ketikan) aksara sunda dengan beberapa macam *font* yang berbeda. Data pada Tugas Akhir ini memiliki ekstensi file .jpg atau .png

3. Perancangan Sistem

Pada tahap perancangan sistem akan dibangun 2 skema, yaitu skema latihan dan skema uji. Setiap skema akan diilustrasikan dengan diagram alur sehingga dapat tergambar alur data dan proses apa yang akan dieksekusi selanjutnya.

4. Implementasi Sistem

Pada tahap ini akan dilakukan implementasi sistem sesuai dengan perancangan yang telah dilakukan pada tahap sebelumnya. Implementasi Tugas Akhir ini menggunakan bahasa Python dengan IDE Jupyter Notebook.

5. Pengujian dan Analisis

Pada tahap ini dilakukan pengujian dengan menggunakan dataset yang telah dikumpulkan terhadap sistem yang telah dibangun. Kemudian hasil dari pengujian tersebut akan dianalisis untuk mengetahui bagaimana performansi sistem serta untuk mendapatkan parameter-parameter terbaik pada sistem.

6. Penyusunan Laporan TA

Pada tahap ini dilakukan penyusunan laporan TA berupa dokumentasi setiap alur metodologi dari tahap studi literatur hingga tahap pengujian dan analisis.

1. 6 Jadwal Kegiatan

Jadwal kegiatan pengerjaan Tugas Akhir ini akan disesuaikan dengan Tabel 1.1

Tabel 1 Jadwal Kegiatan

| No | Kegiatan | Bulan Ke- | | | | | |
|----|------------------------|-----------|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 |
| 1 | Studi Literatur | | | | | | |
| 2 | Pengumpulan Data | | | | | | |
| 3 | Perancangan Sistem | | | | | | |
| 4 | Implementasi Sistem | | | | | | |
| 5 | Pengujian Dan Analisis | | | | | | |
| 6 | Penulisan Laporan TA | | | | | | |