

BAB I

PENDAHULUAN

1.1 Latar Belakang

Indonesia merupakan Negara yang kaya akan kebudayaan. Disetiap daerah pasti memiliki caranya masing-masing dalam mengekspresikan produk kebudayaannya. Baik dalam bentuk bangunan, tarian, bahasa dan juga tulisan. Seperti Rumah Gadang yang merupakan rumah adat Minangkabau, kesenian Reog dari daerah Jawa Timur, bahasa Sunda dari tanah Jawa Barat, dan kebudayaan-kebudayaan dari daerah lainnya.

Jawa Barat merupakan daerah di Indonesia yang memiliki kekayaan peninggalan masa lalu. Salah satunya adalah penggunaan sistem ejaan suatu bahasa atau disebut ortografi. Di Jawa Barat ini terdapat beberapa jenis ortografi yaitu salah satunya aksara sunda. Aksara sunda merupakan salah satu aksara tradisi dari masyarakat Sunda yang sudah ada sejak 5 abad yang lalu hingga saat ini (Baidillah dkk, 2008). Untuk itu dibutuhkan peningkatan upaya pelestarian karena secara konseptual bahasa akan bertahan jika memiliki sistem penulisan atau aksara sebagai fasilitas untuk merekam bahasa itu sendiri (Endang Turmudi, 2013).

Beberapa masalah yang timbul adalah hanya beberapa kalangan yang dapat membaca aksara sunda. Padahal berbicara tentang aksara sunda, banyak naskah kuno yang menggunakan aksara ini sebagai sistem penulisan. Dari naskah-naskah kuno tersebut banyak informasi yang dapat digali, karena naskah kuno memiliki nilai informasi yang sangat berharga baik ditinjau dari aspek sejarah naskah tersebut maupun kandungan informasi yang termuat di dalam naskah tersebut (Yona Primadesi, 2010). Salah satu cara yang dapat dilakukan untuk membantu membacanya adalah dengan menggunakan dukungan teknologi. Teknologi yang dimaksud adalah *Optical Character Recognition* (OCR) atau pengenalan karakter. OCR merupakan perangkat lunak yang dapat mengenali karakter-karakter dari dokumen teks yang berbentuk citra.

Penelitian tentang OCR pada aksara sunda sebelumnya sudah pernah dilakukan. Contohnya adalah penelitian dari Achmad Ridwan mengenai

Pengenalan Tulisan Tangan Aksara Sunda Ngalagena Dengan Algoritma Fuzzy C-Means serta penelitian dari Mubarak yaitu Pengenalan Tulisan Tangan Aksara Sunda Menggunakan Kohonen Neural Network. Dari kedua penelitian tersebut dilakukan OCR dengan *input* berupa karakter tunggal dari aksara sunda. Hasilnya pun sudah cukup baik dengan tingkat akurasi masing-masing untuk penelitian oleh Achmad Ridwan sebesar 70% (Ridwan, 2011) dan penelitian oleh Mubarak sebesar 75,36% (Mubarak, 2010). Dari penelitian yang dilakukan Mubarak, disarankan agar dilakukan penelitian tentang OCR dengan *input* berupa kata atau kalimat dikarenakan memiliki tingkat kesulitan yang lebih tinggi dibanding dengan *input* karakter tunggal.

Dari saran yang telah disebutkan oleh Mubarak, maka timbul masalah baru. Karena jika penelitian tentang OCR tersebut dilakukan, *input*-nya adalah berupa kata atau kalimat, sedangkan dalam pengenalan karakter *input*-nya harus berupa karakter tunggal. Maka dari itu diperlukan pra-proses terlebih dahulu yang dapat membagi *input*-an yang berupa citra berisi kata atau kalimat menjadi sekumpulan citra dengan isi karakter tunggal. Kesulitan lainnya adalah terdapatnya beberapa karakter aksara sunda yang memiliki spasi di dalam karakter itu sendiri. Hal tersebut dapat mengakibatkan apakah karakter tersebut akan tetap menjadi satu karakter atau terpisah menjadi beberapa karakter yang berbeda setelah dilakukan pra-proses.

Selain masalah pemisahan antar karakter, salah satu faktor penghambat lainnya dari pengenalan karakter dalam mendapatkan akurasi *output* yang tinggi adalah kualitas dari citra yang di-*input*-kan. Ketika pengenalan karakter dilakukan himpunan piksel akan mudah disalahartikan jika binarisasi label belum berhasil memisahkan kluster. Salah pengertian lainnya dapat terjadi dengan mudah jika kluster salah terbagi (Sauvola, 1999). Semakin baik kualitas dari citra yang di-*input*-kan, semakin tinggi pula tingkat akurasi yang dihasilkan. Selain itu faktor pencahayaan yang kurang merata pada citra juga dapat mengurangi kualitasnya. Maka dari itu dilakukan *image processing* terlebih dahulu untuk meningkatkan kualitas citra.

Metode yang akan digunakan adalah metode Sauvola. Metode tersebut dilakukan pada tahap *thresholding* untuk mengantisipasi ketidakmerataannya pencahayaan. Metode ini merupakan salah satu metode *adaptive thresholding* dan lebih menghususkan untuk citra dokumen. Degradasi seperti pada pencahayaan dan terdapatnya *noise* dikelola dalam setiap struktur metode ini untuk secara efektif menyaring ketidaksempurnaan. Metode ini pun memiliki pilihan untuk mempercepat proses *thresholding* dengan cara menghitung untuk setiap piksel dan menginterpolasi nilai *threshold* piksel-piksel diantaranya. Pada tahun 1998, metode ini diaplikasikan pada database besar dari citra dokumen yang memiliki 15 perbedaan jenis dokumen. Hasilnya metode *Sauvola* berada di peringkat pertama dengan nilai 94.9% diikuti Niblack (93.7%), Eikvil(82.2%), Bernsen(79.7%), Parker(75.4%) (Sauvola, 1998).

Sebelum masuk ke dalam proses OCR, citra yang berisi kata atau kalimat menggunakan karakter sunda akan dipotong-potong per karakter terlebih dahulu. Maka dibutuhkan sebuah metode yang dapat memisahkan antar karakter dari sebuah kata atau kalimat. Peneliti menggunakan metode *Projection Profile* untuk melakukan pemisahan tersebut. Metode ini dipilih karena mampu melakukan pemisahan karakter, kata, serta baris dan diharapkan mampu menyegmentasi yang memiliki susunan karakter yang berbaris-baris. Penelitian serupa telah dilakukan oleh Ni Putu Sutramiani dengan menggunakan *Projection Profile* pada citra lontar berisi aksara bali, dan menyimpulkan hasil segmentasi mendekati baik dengan rata-rata nilai 2,68 dari 4 (Ni Putu Sutramiani, 2015). *Projection profile* bekerja dengan cara menjumlahkan nilai 1 dari setiap piksel pada citra setelah dilakukan *thresholding*.

Berdasarkan pemikiran diatas, maka dilakukan penelitian mengenai penerapan Sauvola dan *Projection Profile* pada segmentasi karakter aksara sunda. Dengan menggunakan kombinasi kedua metode tersebut, diharapkan hasil segmentasi aksara sunda yang dihasilkan memiliki tingkat akurasi yang tinggi, sehingga dapat menjadi *input* yang baik untuk OCR.

1.2 Rumusan Masalah

Penelitian penggunaan metode *Sauvola* dan *Projection Profile* dalam segmentasi aksara sunda dilakukan untuk mengetahui:

- a. Bagaimana hasil citra biner dan penerapan metode *Sauvola* dalam proses *thresholding* pada segmentasi aksara sunda?
- b. Bagaimana hasil citra segmentasi dan penerapan metode *Projection Profile* yang telah dimodifikasi dalam segmentasi aksara sunda?
- c. Bagaimana data tingkat akurasi yang dihasilkan dari segmentasi citra aksara sunda?

1.3 Tujuan

Tujuan dari penelitian yang dilakukan dalam segmentasi aksara sunda dengan menggunakan metode *Sauvola* dan *Projection Profile* adalah sebagai berikut:

- a. Mendapatkan citra biner yang *background* dan *foreground*-nya terpisah dengan baik.
- b. Mendapatkan citra segmentasi dari penerapan metode *Projection Profile* yang telah dimodifikasi dalam segmentasi aksara sunda.
- c. Mendapatkan data tingkat akurasi dari segmentasi aksara sunda.

1.4 Manfaat Penelitian

Beberapa manfaat yang dapat diperoleh dari penelitian ini adalah sebagai berikut:

- a. Ikut serta dalam melestarikan budaya yaitu aksara sunda.
- b. Mempermudah dalam mengenali aksara sunda.
- c. Menjadi penghasil *dataset* untuk *Character Recognition*.
- d. Menambah pengetahuan tentang seluk beluk aksara sunda, metode *Sauvola* dan *Projection Profile* bagi peneliti khususnya serta bagi pembaca umumnya.

1.5 Batasan Masalah

Adapun batasan masalah dalam penelitian ini adalah:

- a. Masukan yang diterima komputer bersifat *offline*. Yaitu dengan cara mengunggah *file* citra berisi kata atau kalimat dengan menggunakan aksara sunda.
- b. *File* citra yang akan di ujikan merupakan hasil dari foto menggunakan kamera *smartphone*.
- c. Karakter-karakter yang digunakan adalah aksara *ngalagena*, aksara *swara*, dan angka.
- d. Aplikasi yang dibuat ditulis dengan menggunakan Bahasa R sebagai bahasa pemrograman.

1.6 Sistematika Penulisan

Sistematika penulisan yang akan digunakan pada penelitian ini diantaranya sebagai berikut :

BAB I PENDAHULUAN

Bab I atau pendahuluan ini menjelaskan tentang latar belakang mengapa penelitian ini dilakukan. Lalu dijelaskan juga rumusan masalah yang menjadi fokus penelitian, tujuan yang ingin dicapai, manfaat dari dilakukannya penelitian, batasan masalah dan sistematika penulisan.

BAB II KAJIAN PUSTAKA

Pada bab ini diuraikan landasan-landasan teori yang berhubungan dengan penelitian untuk menjadi acuan yang memberi penjelasan permasalahan. Seperti menjelaskan tentang teori metode *Sauvola*, metode *Projection Profile*, dan penjelasan karakter-karakter pada aksara sunda.

BAB III METODE PENELITIAN

Bab III berisi metode yang digunakan dalam melakukan penelitian meliputi desain penelitian, serta alat dan bahan yang dibutuhkan. Bagian-bagian tersebut akan dijelaskan pada bab ini.

BAB IV PEMBAHASAN

Bab ini berisi tahapan-tahapan yang dilakukan pada penelitian dan memaparkan hasil serta analisisnya yang merupakan jawaban dari rumusan masalah penelitian ini.

BAB V KESIMPULAN

Bab V berisi tentang kesimpulan dan saran. Kesimpulan merupakan rangkuman dari hasil penelitian yang telah dilakukan. Serta saran yang memberikan rekomendasi untuk penelitian selanjutnya.