

TRANSLITERATION LAMPUNG SCRIPT WITH METHOD INTENSITY OF CHARACTER

Nanang Himawan Fauzi, Evannoah Rolimarch Pratama, Reynaldi Setiawan, Yakobus Aris Arvanto, Yoga Dwi Prasetyo

Program Studi Teknik Informatika, Universitas Sanata Dharma
Paingan, Maguwoharjo, Depok, Sleman, Daerah Istimewa Yogyakarta 55281
Email : nanang1himawan@gmail.com

Abstract

The goal of writing this scientific work is to explain one of the ways to read Lampung Script using pattern recognition method. There are three main steps in lampungnese image transliteration, i.e. step to preprocessing, feature extracting, and classifying. Each character is made into a binary image that is as thin as possible and the same size. The feature of every character is extracted using intensity of character. For the classification of this pattern recognition is using 1-Nearest Neighbor. By using 120 data of Lampung's script in 6 different font, we got the accuracy of success 95%. From this accuracy we consider that the experiment is a success.

Keyword : Lampung script, pattern recognition, K-Nearest Neighbor, binerization, Rosenfeld

1. Pendahuluan

Salah satu kekayaan budaya Indonesia yang hampir punah adalah aksara Lampung, penggunaannya semakin jarang dan lingkup pemakaiannya semakin sempit. Akibatnya, banyak pustaka leluhur semakin ditinggalkan, termasuk aksara Lampung menjadi kian terpinggirkan dan rapuh, tidak dapat berkembang, dan menjadi sulit mengakomodasikan pemikiran - pemikiran baru menggunakan aksara Lampung.

Sudah banyak usaha yang dilakukan, termasuk pengenalan aksara Lampung dalam kurikulum sekolah tingkat dasar, sehingga menjadi pondasi bagi pelestariannya. Namun demikian, dunia pendidikan memerlukan alat bantu dalam penerbitan literatur yang bermutu. Pada tingkat selanjutnya, para pakar budaya dan bahasa memerlukan alat bantu untuk mengkomunikasikan hasil kajiannya. Semakin canggih alat bantu, semakin terjamin kelancaran dan keruntutan roda pendidikan dan pendalaman budaya Lampung ke masa depan. Salah satu alat bantu tersebut adalah komputer, yang dapat dipergunakan untuk melakukan alih aksara secara otomatis citra aksara Lampung.

Terdapat banyak penelitian yang sudah dilakukan dan terkait proses transliterasi secara keseluruhan, maupun bagian-bagiannya, termasuk penelitian di bidang pengenalan pola, yang dapat digunakan hasilnya untuk kepentingan pengenalan aksara Lampung. Pengenalan aksara ini menjadi hal yang sangat penting dalam proses alih aksara citra dokumen yang ditulis dengan aksara apapun, karena setelah semua citra aksara penyusun dokumen dapat diambil, maka langkah utama yang berikutnya adalah alih aksara citra aksara yang telah terambil.

Di Indonesia sendiri, penelitian terkait alih aksara Jawa maupun transliterasi dokumen beraksara daerah sudah banyak dilakukan. Misalnya Widiarti [1] telah menyimpulkan bahwa metode penipisan citra aksara Jawa Rosenfeld memberikan kinerja yang terbaik dibandingkan dengan metode Hilditch, Zhang-Suen, dan Nagendraprasad-Wang-Gupta. Dengan ditipiskan dulu, maka representasi data citra aksara Jawa menjadi lebih sederhana dan diharapkan akan mempercepat proses alih aksara Jawa. Mempergunakan ciri jumlah intensitas warna hitam pada citra biner suatu aksara [2], Widiarti dan Harjoko [3] mendapatkan persentase

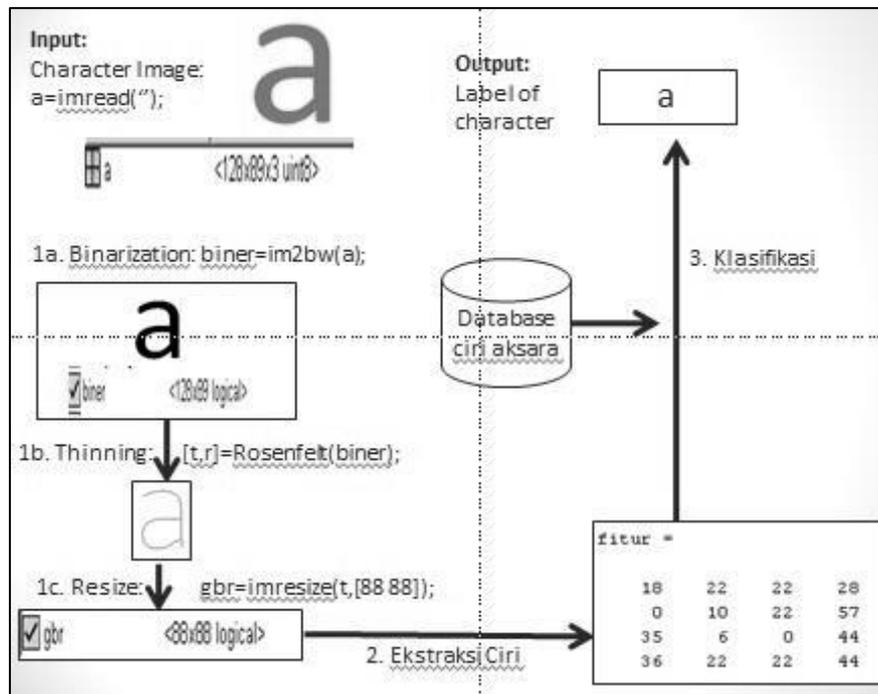
pengenalan sebesar 81,28% pada studi kasus di 4 halaman dari dua buku sastra Jawa modern "Menak Sorangan I", dan Widiarti, dkk. [4] mendapatkan prosentase pengenalan lebih dari 75% untuk mengenali aksara Jawa tulisan tangan pada suatu manuskrip. Paper ini menyajikan hasil penelitian untuk alih aksara Lampung dengan mempergunakan pendekatan yang telah dilakukan untuk alih aksara Jawa tersebut di atas [3][4]. Melalui penelitian ini, diharapkan mesin utama untuk alih aksara Lampung, bisa dibuatkan semacam aplikasi untuk pengenalan aksara lampung pada anak-anak yang berbasis suara. Karena berdasarkan penelitian Wintolo dan Widiastuti (2015), anak-anak terbantu sekali dalam mencoba mengenali sesuatu menggunakan suara, contohnya pengenalan warna [8].

2. Metodologi Penelitian

Surinto [2] menyodorkan berbagai macam ekstraksi ciri untuk pengenalan aksara Thailand, salah satunya adalah dengan menghitung jumlah piksel hitam atau *intensity of character (IoT)* penyusun citra aksara biner yang telah dibagi luasannya menjadi 9 bagian, seperti terlihat pada rumus (1). Jika terdapat citra S yang dibagi menjadi q baris dan r kolom, maka jumlah piksel hitam pada luasan tersebut dapat dihitung dengan rumus (1). Cara tersebut lalu diadopsi peneliti lain untuk alih aksara Jawa [3][4], dan kemudian dipergunakan dalam penelitian ini.

$$cr_{q,r} = \sum_{x=(q-1)unit+1}^{q.unit} \sum_{y=(r-1)unit+1}^{r.unit} S_{x,y}, \quad q, r = 1.. \frac{n}{unit} \tag{1}$$

Dalam Gambar 1 diperlihatkan garis besar metode pengembangan alat uji untuk memeriksa kinerja IoT dalam alih aksara Lampung. Sebelum citra aksara diproses, sebelumnya dilakukan proses untuk binerisasi citra dengan mempergunakan metode Otsu yang telah diimplementasikan dalam modul `im2bw()` di Matlab.

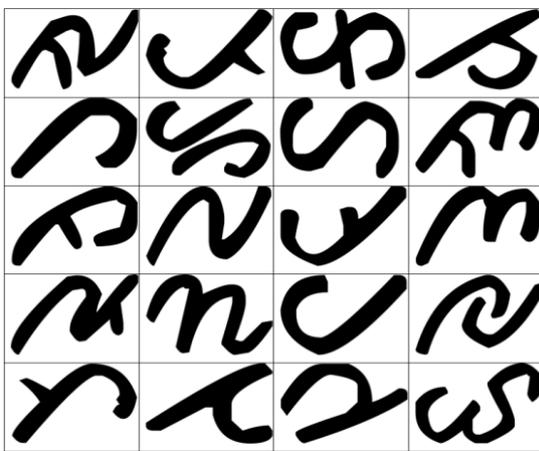


Gambar 1. Model alih aksara yang diadopsi dalam penelitian [5]

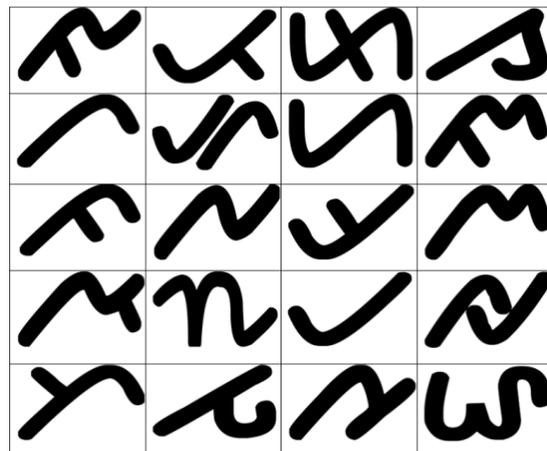
Setelah menjadi citra biner, dilanjutkan dengan penipisan citra aksara untuk mempercepat proses komputasi serta untuk mendapatkan bentuk kerangka dasar setiap citra aksara dengan mempergunakan metode penipisan Rosenfeld [1]. Setelah citra menjadi lebih tipis pada setiap

kerangka pembentuknya, untuk menjamin bahwa untuk setiap citra aksara yang menyimbolkan aksara sama memungkinkan mempunyai ciri yang sama, maka dilakukan penyamaan ukuran citra menjadi 88x88 piksel sebelum diekstraksi ciri. Pada tahap klasifikasi, penelitian ini mempergunakan metode pencocokan kemiripan *1-Nearest Neighbour* saja karena metode ini populer dan data yang dipergunakan adalah citra aksara Lampung cetak. Untuk mengukur kedekatan citra latih dan citra uji dipergunakan rumus jarak *City Block*[6].

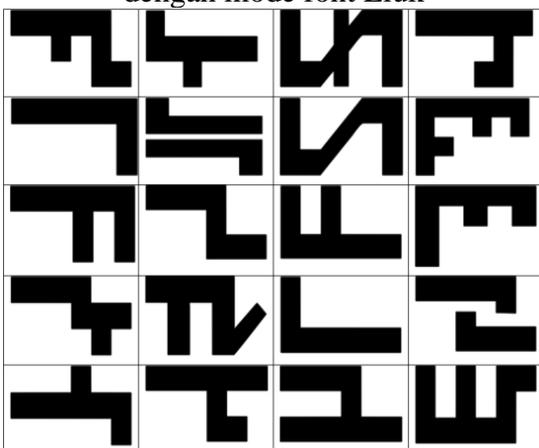
Jumlah data yang dipergunakan sebanyak 120 citra aksara Lampung dalam 6 font berbeda untuk 20 aksara Lampung yang berbeda. Dengan mempergunakan program MsWord yang sudah diinstal dengan font Digital, Liuk, YZ, YZ Rounded, YZ Hand, dan YZ. , ke-120 aksara tersebut diketik. Setiap citra aksara tersebut kemudian di-*capture* menggunakan *snipping tool* yang menghasilkan gambar dengan ekstensi PNG. Menggunakan cara berpikir Kusumaningrum[7], maka data untuk training dan testing dibuat dengan cara menuliskan aksara Lampung dalam 6 font berbeda. Di Gambar 2 sampai Gambar 7 diperlihatkan sebagian contoh sampel citra aksara Lampung dalam font yang menjadi data uji dalam penelitian ini.



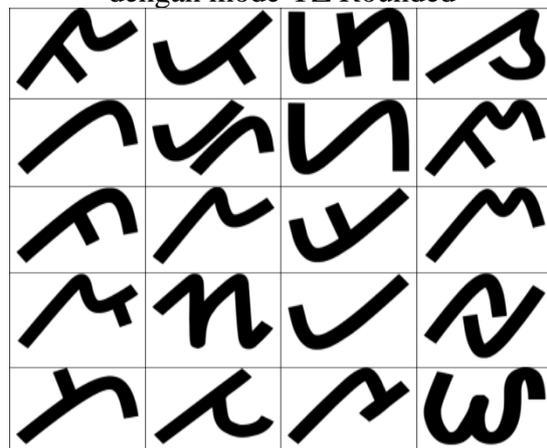
Gambar 2. Contoh citra aksara Lampung dengan mode font Liuk



Gambar 3. Contoh citra aksara Lampung dengan mode YZ Rounded



Gambar 4. Contoh citra aksara Lampung dengan mode font Digital



Gambar 5. Contoh citra aksara Lampung dengan mode YZ.

Dari 40 data uji dalam percobaan, diperoleh 38 data yang dikenali dengan benar dan 2 data lainnya salah dikenali. Maka persentase kebenaran yang diperoleh sebesar 95%, sehingga dapat disimpulkan bahwa metode ekstraksi ciri yang dipilih sudah sesuai dengan karakteristik aksara Lampung. Tempat proses terjadinya kesalahan dari 2 data alih aksara tersebut terjadi pada saat melakukan pengukuran kedekatan citra uji ke citra latih dengan menggunakan rumus *City Block*. Kesalahan ini juga dikarenakan pada percobaan ini hanya menggunakan metode pencocokan *1-Nearest Neighbour*, apabila percobaan menggunakan metode pencocokan *k-Nearest Neighbour* dengan k lebih besar dari 1 maka persentase kebenaran yang diperoleh bisa menjadi lebih besar. Kesalahan terjadi pada alih aksara BA yang dikenali sebagai PA pada font YZ. , dan aksara LA yang dikenali sebagai WA pada font Liuk, karena memang baik BA dan PA, maupun LA dan WA tingkat kemiripannya besar.

4. Kesimpulan

IoT sebagai pencari pada aksara Lampung cetak telah menghasilkan tingkat pengenalan yang sangat besar. Dengan klasifikasi *1-Nearest Neighbour*, ciri tersebut bahkan mampu mengenali aksara Lampung dengan baik. Dengan adanya beberapa aksara Lampung yang mirip, misal aksara BA dengan PA, aksara LA dengan aksara WA, membuka kemungkinan untuk penelitian lebih lanjut agar aksara-aksara Lampung yang mempunyai kemiripan besar dapat dibedakan dengan lebih baik lagi.

Ucapan Terimakasih

Terima kasih kepada Universitas Sanata Dharma (USD) yang memberikan bantuan dana sehingga penelitian dan deseminasi hasil penelitian bersama mahasiswa Teknik Informatika USD dapat dilaksanakan.

Daftar Pustaka

- [1] Widiarti, A.R. (2011). Comparing Hilditch, Rosenfeld, Zhang-Suen, and Nagendraprasad-Wang-Gupta Thinning Algorithms for Javanese Character Image. In *International Conf. on Pattern Recognition and Computer Vision (ICPRCV)*. Amsterdam Netherlands(pp. 938–942).
- [2] Surinta, O. (2010). *Overview of Handwritten Thai Character Recognition*. online: <http://www.ai.rug.nl/~mrolarik/APSMeeting/09-072010%20Overview%20of%20Handwritten%20Diakses>
- [3] Widiarti, A.R., dan Harjoko, A. (2006). Pengenalan Citra Dokumen Sastra Jawa (Studi Kasus Pada Buku Sastra Jawa: Menak Sorangan I dan Panji Sekar). *Sains Dan Sibernetika*, 19(3, pp. 269–282)
- [4] Widiarti, A. R., Harjoko, A., Marsono, dan Hartati, S. (2017). The Model and Implementation of Javanese Script Image Transliteration. In *International Conference on Soft Computing, Intelligent System and Information Technology (ICSIT)* Denpasar(pp. 51-57).
- [5] Widiarti, A.R. (2015). *Pengenalan Aksara. Studi Kasus Pengenalan Aksara Jawa*. online: <http://exelsa.usd.ac.id/modules.php?mod=2e66d75ab544cd2d72605aa84a4100-e4&pid=5314> Diakses 2 Maret 2018.
- [6] Cha, S-H. (2007). Comprehensive Survey on Distance/Similarity Measures between Probability Density Functions. *International Journal Of Mathematical Models And Methods In Applied Sciences*, 4(1, pp:300-307).
- [7] Kusumaningrum, A. (2017). Pengaruh Jumlah file Training Terhadap Akurasi Pendeteksian Obyek Pada Metode Viola Jones. In *Seminar Nasional Teknologi Informasi dan Kedirgantaraan (SENATIK) 2017*, 3(pp:1).

- [8] Wintolo H., dan Widiastuti, T.W. (2015). Pemanfaatan Image To Speech Berbasis Android Untuk Pengenalan Warna Bagi Anak Bawah Tiga Tahun (Batita). In *Jurnal Ilmiah Bidang Teknologi Angkasa*, 7(pp:2).
- [9] Rosandy, T. (2016). *Perbandingan Metode Naive Bayes Classifier Dengan Metode Decision Tree (C4.5) untuk Menganalisa Kelancaran Pembiayaan* ,2(pp;57-58).
- [10] J.L. Danhauer and L.E. Lucks. (1987). *The Confusion Matrix: A New Model*. In *Human Communication Canada/Communication Humaine Canada, Vol. 11, No. 4, 1987*(pp;2).