
PARTISI BLOK TEKS MENUJU RESTORASI KERUSAKAN AKSARA JAWA

¹ Agustinus Rudatyo Himamunanto, ² Emeritas Setyowati
^{1,2} Fakultas Sains dan Komputer, Universitas Kristen Immanuel
Jalan Solo Km. 11 PO Box 4 YKAP Yogyakarta, ph: (0274) 496256-296247 fax: (0274) 496258
e-mail: rudatyo@gmail.com, emeritas@gmail.com

Abstrak

Pada umumnya sebuah lembar dokumen terdapat dua obyek informasi yaitu teks dan gambar. Blok teks merupakan obyek vital yang ada dalam sebuah lembar dokumen karena hanya pada area teks inilah tahapan proses restorasi akan dilakukan. Operasi identifikasi blok teks dari obyek gambar menjadi tahapan penting. Paper ini memaparkan satu metode identifikasi area blok teks aksara Jawa melalui karakteristik aksara Jawa serta cara penulisannya. Berdasarkan pengujian pada input naskah Hamong Tani diketahui bahwa system bekerja dengan tingkat keberhasilan 93,5%. Area teks yang belum teridentifikasi merupakan cap stempel perpustakaan itupun karena dalam posisi miring.

Kata kunci : identifikasi, blok teks, aksara Jawa

Abstract

Generally a sheet of documents there are two objects of information, namely text and images. A text block is a vital object in a sheet of documents because it is only in this text area that the stage of the restoration process will take place. The identification operation of the text block from the image object becomes an important step. This paper describes a method of identifying areas of Javanese script block through the characteristics of Javanese script as well as the way of writing. Based on the test on the input manuscript Hamong Tani known that the system works with a success rate of 93.5%. Unidentified text area is the seal of the library stamp because it is in an oblique position.

Keywords : identification, text block, Javanese script

1. LATAR BELAKANG

Manuskrip Jawa adalah salah satu warisan karya sastra Jawa yang banyak bercerita tentang sejarah perkembangan kehidupan ditanah Jawa. Manuskrip ini ditulis baik dalam aksara Jawa tulisan tangan maupun aksara Jawa cetak. Menurut Anastasia Rita W. [9] kondisi sebuah manuskrip yang baik adalah jika manuskrip tersebut masih memiliki kelengkapan informasi visual yang memadai pada setiap karakternya dimana pada setiap bentuk karakternya masih tetap utuh, jelas dan benar. Pada sisi yang lain, manuskrip adalah salah satu warisan budaya yang paling mudah mengalami kerusakan yang diakibatkan baik oleh alam maupun kecerobohan manusia sendiri. Kerusakan visual pada manuskrip menyebabkan makna yang

tersurat dalam manuskrip tersebut menjadi tidak mudah dipahami. Diperlukan seorang ahli sastra Jawa untuk melakukan tahapan restorasi manual atau perbaikan kerusakan informasi visual pada dokumen manuskrip yang mengalami kerusakan.

Proses restorasi manual pada kenyataannya memerlukan waktu yang cukup lama serta membutuhkan kesabaran yang tinggi. Namun begitu, O’Gorman L dan Kasturi R.[7] menyampaikan bahwa perkembangan bidang teknologi pengolahan citra digital berpeluang memberikan kesempatan untuk membantu mempercepat berbagai pekerjaan termasuk didalamnya adalah proses perbaikan kerusakan obyek karakter, seperti pada proses restorasi aksara Jawa. Perkembangan sekecil apapun pada model proses restorasi digital aksara Jawa diharapkan memberi peluang selanjutnya bagi peningkatan teknologi restorasi digital pada kerusakan visual aksara Jawa cetak.

2. MODEL DOKUMEN AKSARA JAWA

Manuskrip adalah karya sastra kuno yang ditulis dalam berbagai media seperti: daun lontar, batang bambu, kertas, kulit kambing dan lain-lain. Sedemikian banyaknya warisan dokumen sastra di nusantara membuat pekerjaan transliterasi secara manual seolah merupakan proses abadi. Upaya para peneliti bidang teknologi informasi dan komputer memberikan perkembangan dalam memberikan support kepada para ahli sastra jawa dalam melakukan proses transliterasi manuskrip. Proses transliterasi secara manual yang dilakukan oleh para ahli filologi bahasa jawa dipandang sudah mampu mengatasi segala kondisi format lembar dokumen sastra jawa. Namun begitu tidak demikian halnya dengan proses transliterasi otomatis. Saat sebuah citra dokumen diterima oleh sistem transliterasi, salah satu hal dasar yang harus dilakukan adalah memastikan bahwa input transliterasi adalah murni berupa karakter aksara jawa. Pada kenyataannya tidak pada semua lembar input citra dokumen sastra jawa berisi melulu hanya berupa blok teks dan ini berarti perlu dipersiapkan tahapan proses untuk memisahkan komponen blok teks dan blok ilustrasi gambar.

Berikut adalah contoh lembar dokumen sastra jawa kuno terbitan tahun 1876, halaman 56, karangan W.S. Winter dengan judul "Hamong Tani".



Gambar 1. Citra sebuah dokumen aksara Jawa

3. MODEL KERUSAKAN AKSARA JAWA

Beberapa manuskrip terpaksa harus di digitalisasi dalam kondisi sudah mengalami berbagai kerusakan, mulai dari tinta yang sudah samar, sebagian garis karakter yang sudah kabur hingga sebagian media yang sudah sobek dan hilang. Salah satu contoh kerusakan visual adalah distorsi

garis atau menipisnya garis yang membentuk karakter yang diakibatkan oleh hilangnya bagian pinggir dari garis atau bahkan mengalami garis karakter yang putus seperti dicontohkan pada Gambar 2.

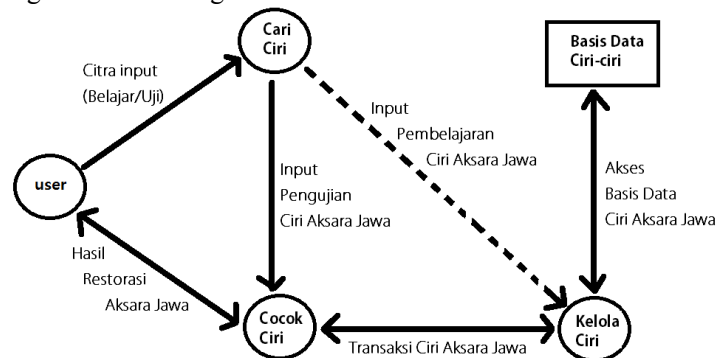


Gambar 2. Contoh 3 model kerusakan visual citra aksara Jawa

Bagi beberapa orang yang sudah cukup terbiasa mempergunakan bahasa Jawa atau setidaknya mengenal aksara Jawa, kerusakan visual citra karakter aksara Jawa boleh jadi tidak memiliki masalah. Permasalahan dimungkinkan muncul manakala pada citra aksara tersebut akan dikenai pemrosesan citra tingkat lanjut, misalnya untuk keperluan transliterasi aksara Jawa [7]. Restorasi citra aksara Jawa menjadi langkah penting agar citra aksara Jawa yang mengalami kerusakan mendapatkan perbaikan visual sedemikian rupa sehingga tidak menimbulkan interpretasi yang bias baik bagi orang yang akan membaca apa yang tersurat maupun bagi pemrosesan lanjut.

4. TAHAPAN PROSES RESTORASI

Proses restorasi digital dalam penelitian ini melibatkan 2 proses utama yaitu proses pembelajaran dan proses pengujian [4][8]. Sebenarnya kedua proses ini memiliki langkah proses yang sama namun dipergunakan untuk kepentingan yang berbeda. Secara umum tahapan proses restorasi digital adalah sebagai berikut:



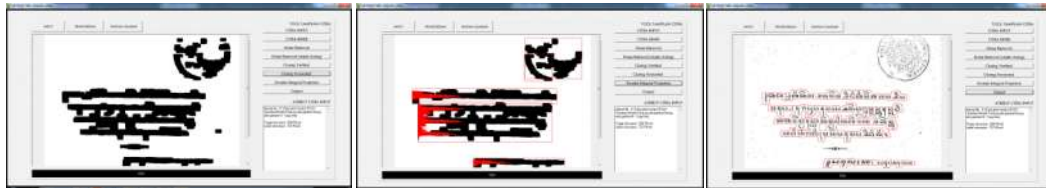
Gambar 3. Diagram proses restorasi pada Kerusakan citra aksara Jawa

5. HASIL PENGUJIAN DAN PEMBAHASAN

Proses pengujian melibatkan input berupa hasil scan dokumen teks aksara Jawa baik yang memiliki gambar ilustrasi maupun yang tidak memiliki gambar ilustrasi. Tahapan pengujian dilakukan secara otomatis yang meliputi langkah umum sebagai berikut: identifikasi blok teks aksara Jawa, segmentasi karakter aksara Jawa, identifikasi kerusakan aksara Jawa, restorasi kerusakan aksara Jawa.

2.1. IDENTIFIKASI BLOKS TEKS (IBT)

Merupakan proses yang dilakukan untuk mengidentifikasi adanya bloks teks aksara Jawa dan mengelompokkan secara terpisah dengan area gambar ilustrasi.



Gambar 4. Hasil proses Identifikasi Blok Teks (IBT)

2.2. SEGMENTASI OBYEK KARAKTER AKSARA JAWA

Input para proses ini berupa hasil dari proses identifikasi blok teks yang sudah dilakukan sebelumnya, sehingga hanya obyek teks saja yang diproses dalam segmentasi obyek karakter aksara Jawa. Proses ini dilakukan untuk mendapatkan secara mandiri karakter demi karakter darai aksara. Adapun metode yang dipergunakan untuk mendapatkan setiap obyek karakter adalah profil proyeksi horizontal dan vertikal. Berikut adalah hasil dari operasi segmentasi obyek karakter aksara Jawa.



Gambar 5. Hasil proses segmentasi karakter aksara Jawa

2.3. RESTORASI KARAKTER AKSARA JAWA

Input pada proses ini berupa hasil dari proses segmentasi karakter aksara Jawayang sudah dilakukan sebelumnya, sehingga hanya sebuah obyek karakter saja yang diproses dalam restorasi karakter aksara Jawa. Proses ini meliputi identifikasi kerusakan untuk mengecek apakah terdapat kerusakan atau tidak pada obyek karakter yang sedang diamati. Adapun metode pembelajaran yang dipergunakan untuk mendapatkan informasi hasil identifikasi dan atau proses restorasi meliputi: operasi penipisan (thinning), operasi penelusuran piksel dengan metode Freeman Chain Code untuk mendapatkan ciri obyek karakter, inialisasi manual obyek karakter serta penyimpanan informasi dalam basis data. setiap obyek karakter adalah profil proyeksi horizontal dan vertikal. Metode pengujian elakukan langkah yang sama dengan metode pembelajaran, hanya saja cirri karakter yang ditemukan bukan disimpan dalam basis data tapi dicocokkan dengan cirri yang tersimpan dalam basis data. Bila tidak ada yang cocok berarti berpeluang karakter tersebut mengalami kerusakan dan perlu direstorasi. Namun apabila cirri uji memiliki salah satu kecocokan maka berarti tidak terdapat kerusakan atau mungkin karakter haya mengalami kerusakan ringan. hasil Beberapa contoh kerusakan adalah sebagai berikut:

Tabel 1. Contoh gambar kerusakan aksara Jawa level 1

Gambar aksara rusak	Nama	Gambar aksara rusak	Nama
	ho		no
	co		ro
	ko		Do

Tabel 2. Contoh gambar kerusakan aksara Jawa level 2

Gambar aksara rusak	Nama	Gambar aksara rusak	Nama
	ho		No
	co		Ro
	ko		Do

Tabel 3. Contoh gambar kerusakan aksara Jawa level 3

Gambar aksara rusak	Nama	Gambar aksara rusak	Nama
	ho		No
	co		Ro
	ko		Do

Hasil tahapan proses pengujian adalah citra aksara terdekat yang diharapkan memberikan peluang sebaik mungkin untuk menjadi citra kandidat atau citra hasil restorasi. Beberapa contoh hasil pengujian adalah seperti disajikan pada tabel berikut ini.

Tabel 4. Contoh hasil pengujian dengan input beragam level kerusakan aksara Jawa

Gambar input uji	Gambar hasil restorasi	Keterangan
		Benar
		Benar
		Salah
		Benar
		Benar
		Benar
		Salah
		Salah

Berdasarkan hasil pengujian dengan jumlah input sebanyak 452 input model aksara Jawa yang mengalami berbagai kerusakan, maka diketahui jumlah aksara yang dinyatakan menemukan hasil restorasi yang benar adalah sebesar:

$$\text{persentase kebenaran} = \frac{\sum \text{citra benar}}{\sum \text{citra dalam dokumen}} \times 100 = \frac{371}{452} \times 100 = 82.07\%$$

2.4. ANALISA RESTORASI

Berdasarkan pengamatan hasil-hasil restorasi terutama yang menunjukkan hasil restorasi yang tidak benar, diprediksi bahwa proses yang salah berpeluang muncul dari proses dilatasi yang kurang optimal sehingga berujung pada proses penelusuran piksel yang menghasilkan informasi yang kurang tepat. Percobaan juga dilakukan dalam model kerusakan yang bersifat terbatas sehingga hasil percobaan masih bersifat sementara.

6. KESIMPULAN

Berdasarkan beberapa kali hasil pengujian yang melibatkan 3 jenis kerusakan, diketahui bahwa keberhasilan proses restorasi dengan tahapan proses diatas adalah sebesar 82,07%. Kekurangberhasilan dimungkinkan terjadi karena proses dilatasi terhadap kerusakan yang kurang menunjukkan keberhasilan yang kemungkinan berakibat pada proses penipisan yang kurang sempurna sehingga hasil penelusuran piksel juga kurang benar. Namun begitu diharapkan sebarang keberhasilan model restorasi ini akan memberi langkah awal pada perkembangan proses restorasi selanjutnya.

UCAPAN TERIMA KASIH

Terima kasih disampaikan kepada pimpinan Kopertis Wilayah V Daerah Istimewa Yogyakarta atas dukungan dana dalam penelitian ini melalui kontrak nomor: 308/B.Pen.KOP V/VII/2015. Penelitian ini juga mendapatkan dana dari hibah kompetisi Penelitian Dosen Pemula dari Kementerian Riset dan Pendidikan Tinggi Republik Indonesia pada tahun 2016

DAFTAR PUSTAKA

- [1] Chrupala, Grzegorz, 2013, Text Segmentation With Character-Level Text Embeddings, Workshop On Deep Learning For Audio, Speech And Language Processing
- [2] Xu Ceng-Yin, 2013, Robust Text Detection in Natural Scene Images, Arxi vol. 03 no 01
- [3] Chucai-Yi, Yingli-Tian, 2011, Text Detection in Natural Scene Images by Stroke Gabor Words, International Conference on Document Analysis and Recognition
- [4] Chaudhury, S., Jindal, M., Roy, S.D., 2014, Model-Guided Segmentasi and Layout Labelling of Document Images Using A Hierarchical Condition Random Field, Dept. of Electrical Engg. IIT Delhi, Haux Khas, New Delhi
- [5] Jung Jin-Lee, 2011, AdaBoost for Text Detection in Natural Scene, International Conference on Document Analysis and Recognition
- [6] Cong Yao, 2013, Detecting Texts of Arbitrary Orientations in Natural Images, IEEE Computer Society Press.
- [7] Shangxuan Tian, 2015, Text Flow: A Unified Text Detection System in Natural Scene Images, IEEE Computer Society Press.
- [8] Widiarti, Anastasia Rita, 2013, Javanese character image segmentation of document image of Hamong Tani, IEEE Explore.
- [9]
- [10] Livingston, J., Santosh, 2013, Text Detection From Documented Image Using Image Segmentation, International Journal of Technology Enhancement and Emerging Engineering Research, Vol. 1, Issue 4.