

Pendeteksian Huruf Jawa pada Naskah Kuno menggunakan Binerisasi Otsu

A. S. Putra
Universitas Islam Indonesia
Yogyakarta, Indonesia
16523228@students.uii.ac.id

I. Muhimmah
Universitas Islam Indonesia
Yogyakarta, Indonesia
985240102@uui.ac.id

Abstrak—Sudah banyak naskah kuno dari Yogyakarta yang sudah didigitalisasikan. Kebanyakan naskah kuno tersebut ditulis menggunakan Aksara Jawa, sehingga banyak orang yang tidak dapat membaca naskah tersebut. Penelitian ini difokuskan untuk mendeteksi dan mensegmentasi aksara-aksara pada citra naskah kuno yang dapat digunakan dan dikembangkan menjadi identifikasi atau pengenalan Aksara Jawa pada penelitian selanjutnya. Proses segmentasi yang digunakan dalam penelitian ini adalah *Otsu's binerization*. Penggunaan *Otsu's binerization* pada segmentasi citra ini berhasil mendeteksi Aksara Jawa.

Kata kunci—naskah kuno; segmentasi teks; Aksara Jawa; *otsu's binarization*

I. PENDAHULUAN

Di Jawa, salah satunya di Daerah Istimewa Yogyakarta, terdapat banyak warisan-warisan budaya yang penting dalam bentuk dokumen atau naskah kuno. Naskah-naskah tersebut merupakan warisan budaya yang perlu dilestarikan karena merupakan salah satu bukti nyata bahwa Indonesia mempunyai budaya.

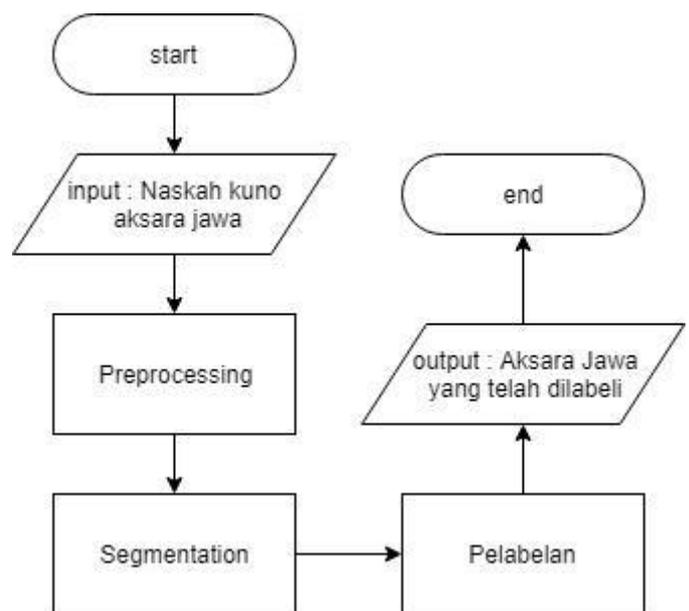
Saat ini, sudah ada 75 naskah kuno dari Yogyakarta yang dimiliki oleh British Library dan 35 diantaranya sudah didigitalisasikan dan akan diserahkan kepada Badan Arsip dan Perpustakaan DIY serta Perpustakaan di Jakarta. Naskah-naskah tersebut juga dapat diakses *online* secara gratis di situs *British Library*. Isi dari suatu dokumen atau naskah tertulis biasanya dapat juga memberikan informasi-informasi dari beberapa aspek budaya bangsa serta masyarakat yang bersangkutan[1].

Naskah-naskah kuno di Yogyakarta kebanyakan ditulis menggunakan karakter atau Aksara Jawa. Permasalahannya informasi-informasi penting pada naskah-naskah tersebut tidak akan sampai ke masyarakat apabila masyarakat tidak bisa membaca Aksara Jawa. Dari permasalahan itu tujuan penulis dari penelitian ini adalah untuk dapat membuat aplikasi yang dapat mendeteksi aksara-aksara Jawa pada naskah Jawa kuno untuk dapat dikembangkan lagi ke arah yang lebih luas seperti pengenalan karakter atau *Optical Character Recognition* (OCR).

Penelitian ini dilakukan untuk mendeteksi tulisan atau aksara Jawa pada citra naskah kuno yang diambil dari *British Library*.

II. METODOLOGI PENELITIAN

A. Gambaran Umum Segmentasi Teks



Gambar 1. Flowchart metodologi penelitian

Diawali dengan proses akuisisi data. Lalu dilanjutkan dengan proses *preprocessing* yang dapat terdiri dari *grayscale* dan *median filter*. Selanjutnya adalah proses segmentasi citra. Pada penelitian ini, proses segmentasi citra menggunakan *Otsu's Binarization*. Setelah itu, proses terakhirnya adalah tahap untuk melabeli aksara menggunakan *connected component labeling*.

B. Akuisisi Data

Akuisisi data merupakan tahap awal dari penelitian ini, yaitu untuk mendapatkan citra digital. Proses ini bertujuan untuk mendapatkan citra digital naskah Kuno yang menggunakan Aksara Jawa. Data-data citra atau gambar yang digunakan dalam penelitian ini diperoleh dari *British Library*.

C. Preprocessing

Proses dari *preprocessing* ini diawali dengan mengubah citra yang berwarna menjadi citra abu-abu atau *grayscale*. Setelah proses *grayscale* selesai, maka dilakukan proses *median filter*. Proses *median filter* ini digunakan untuk mengurangi *noise-noise* atau derau di citra.

D. Segmentasi (Otsu's Binarization)

Setelah selesai proses *preprocessing*, dilanjutkan dengan proses segmentasi. Segmentasi citra itu adalah salah satu bagian penting dari *image processing* untuk membagi-bagi citra atau gambar menjadi beberapa segmen yang sama berdasarkan aturan serta kriteria-kriteria tertentu. Metode segmentasi yang digunakan pada penelitian kali ini adalah *binarization* atau binerisasi. Dalam proses ini citra akan diubah menjadi citra biner atau citra hitam dan putih.

Biasanya, proses binerisasi citra *grayscale* adalah seperti berikut.

$$g(x,y) = \begin{cases} 1 & \text{if } f(x,y) \geq T \\ 0 & \text{if } f(x,y) < T \end{cases} \quad (1)$$

Keterangan :

x adalah baris *pixel* citra

y adalah kolom *pixel* citra

Fungsi $g(x,y)$ adalah citra biner

Fungsi $f(x,y)$ adalah citra *grayscale*

T adalah nilai ambang.

Tetapi metode yang bernama *Otsu's binerization* ini merupakan metode dengan cara membagi histogram gambar *grayscale* menjadi dua daerah yang berbeda secara otomatis tanpa perlu menentukan nilai ambang.

Metode *otsu* menggunakan pendekatan yaitu dengan menggunakan analisis oleh diskriminan dengan menentukan suatu variable yang dapat membeda-bedakan antara dua kelompok atau lebih sehingga akan muncul dengan natura 1. Analisis dari diskriminan *method* ini akan memaksimumkan variabel tersebut supaya bisa membagi objek menjadi latar depan (*foreground*) dan latar belakang (*background*) [2].

Rumus dari *method* binerisasi *otsu's* yaitu sebagai berikut.

Nilai ambang yang akan dicari dari suatu gambar atau citra *grayscale* bisa dinyatakan dengan nilai k . Kisaran dari nilai k adalah antara 1 sampai L , dengan nilai $L = 255$.

Probabilitas dari setiap *pixel-pixel* pada *level* ke- i bisa dinyatakan sebagai berikut:

$$p_i = n_i / N \quad (2)$$

Keterangan :

n_i yaitu menyatakan jumlah *pixel* pada *level* ke i

N merupakan total dari jumlah *pixel* pada citra.

$$\omega(k) = \sum_{i=1}^k p_i \quad (3)$$

$$\mu(k) = \sum_{i=1}^k i.p_i \quad (4)$$

$$\mu_T = \sum_{i=1}^L i.p_i \quad (5)$$

Nilai ambang k bisa ditentukan dengan memaksimumkan dari persamaan :

$$\sigma_B^2(k^*) = \max_{1 \leq k < L} \sigma_B^2(k) \quad (6)$$

dan

$$\sigma_B^2(k) = \frac{[\mu_T \omega(k) - \mu(k)]^2}{\omega(k)[1 - \omega(k)]} \quad (7)$$

E. Pelabelan Aksara

Setelah mendapatkan hasil dari binerisasi, maka dilanjutkan dengan proses untuk melabeli aksara. Proses ini digunakan untuk memberi label yang berbeda pada setiap karakter atau huruf sehingga yang satu dengan yang lain dapat dipisahkan. Proses ini menggunakan metode *connected component labeling*(CCL). Proses ini dapat dilakukan dengan *mengecheck* suatu citra serta mengelompokkannya dari setiap *pixel* ke dalam suatu komponen yang terhubung dengan aturan keterhubungan.

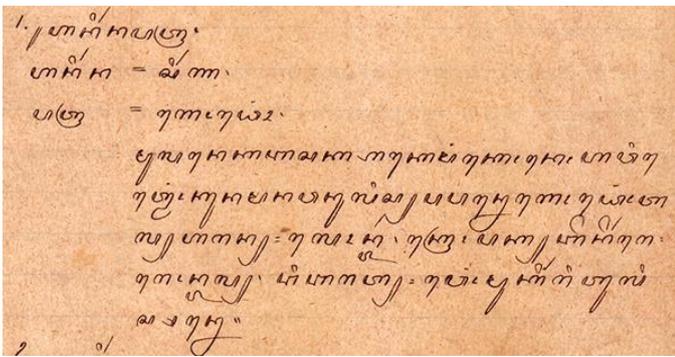
Setiap dari komponen yang terhubung tersebut apabila ada yang saling tidak terhubung maka akan diberi dengan label yang berbeda.

III. PENGUJIAN

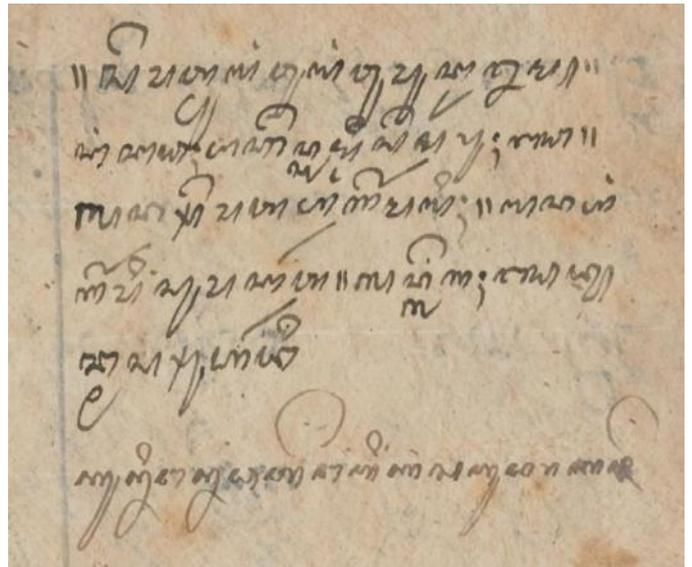
A. Data Uji Coba

Data citra yang dipakai pada penelitian ini terdiri atas 5 citra dari naskah kuno yang diambil dari British Library yang dimana model tulisan dari naskah-naskah tersebut berbeda.

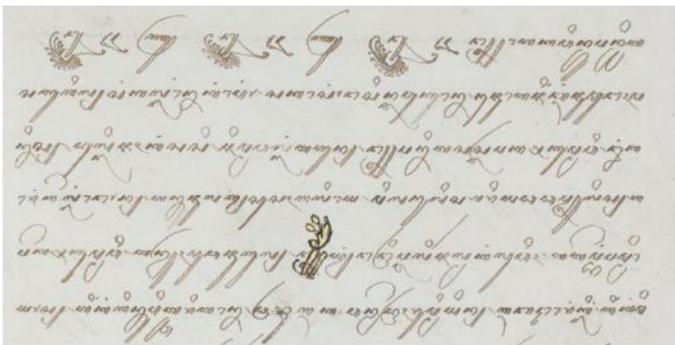
Data citra yang digunakan adalah:



Gambar 2. Data citra uji coba 1



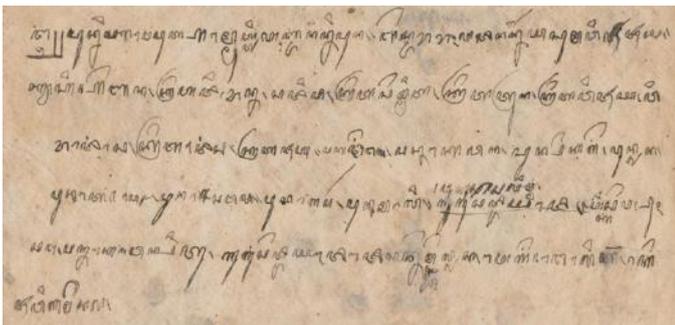
Gambar 6. Data citra uji coba 5



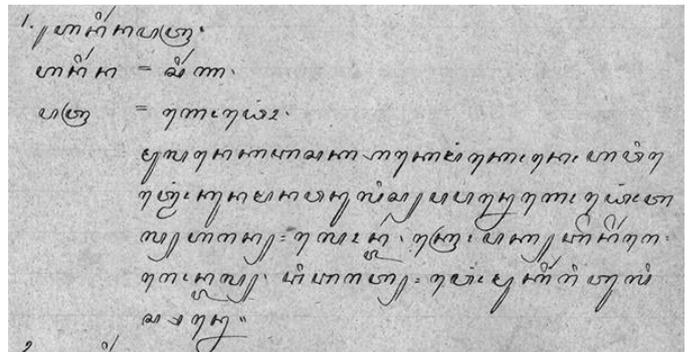
Gambar 3. Data citra uji coba 2

B. *Grayscale*

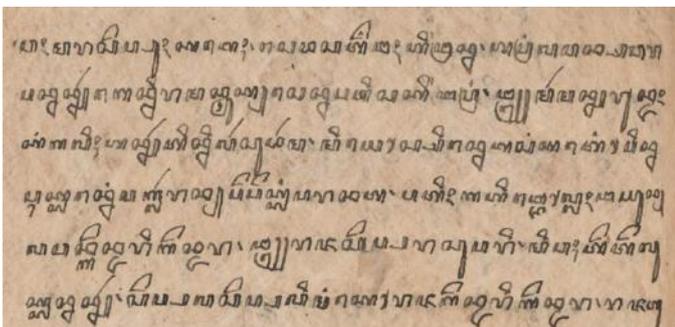
Setelah mendapatkan data citra maka proses yang dilakukan adalah *grayscale*. Hasil dari *Grayscale* adalah citra abu-abu seperti berikut



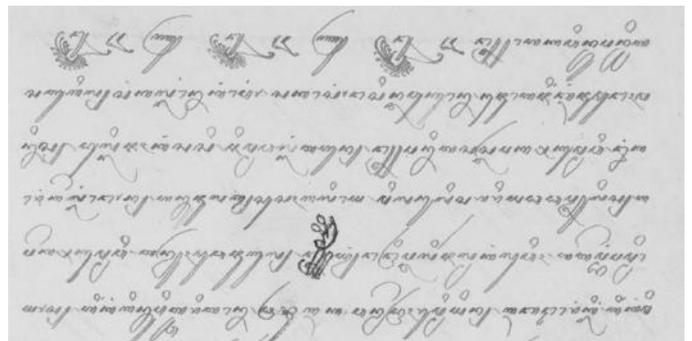
Gambar 4. Data citra uji coba 3



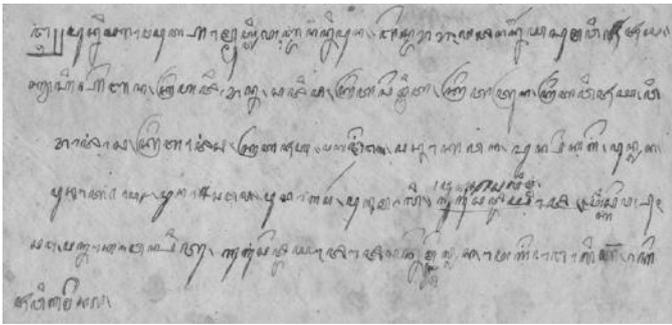
Gambar 7. Citra gray dari gambar uji coba 1



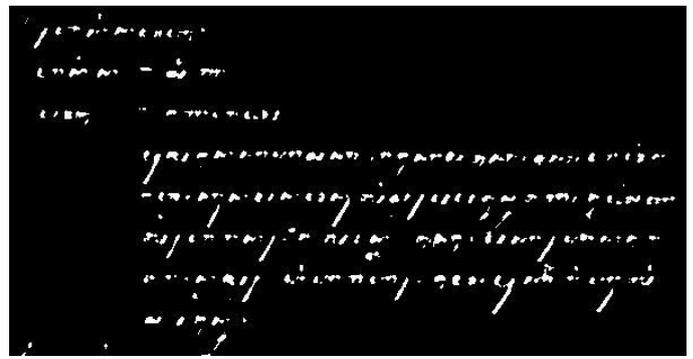
Gambar 5. Data citra uji coba 4



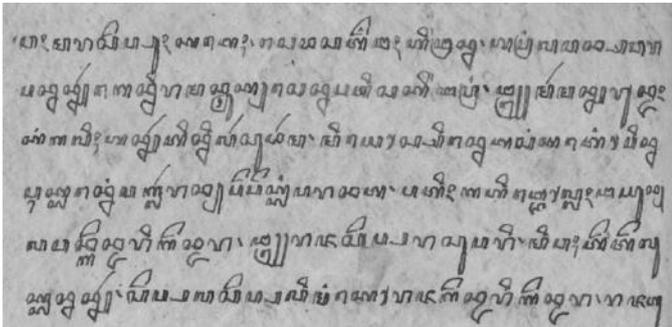
Gambar 8. Citra gray dari gambar uji coba 2



Gambar 9. Citra gray dari gambar uji coba 3



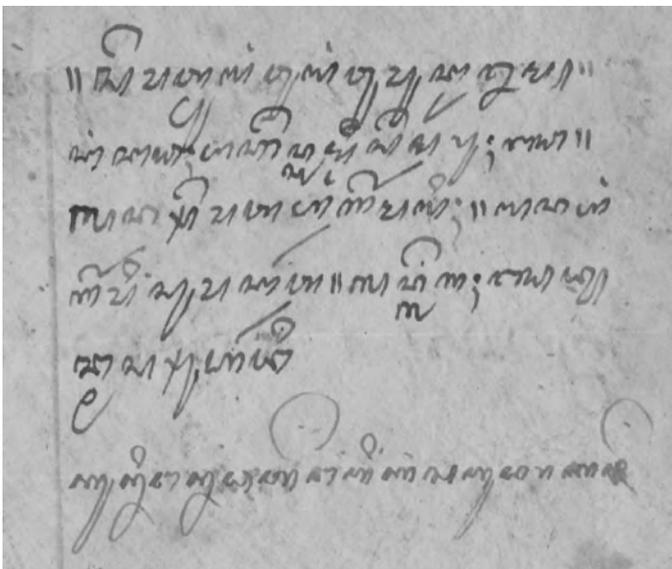
Gambar 12. Citra biner dari gambar uji coba 1



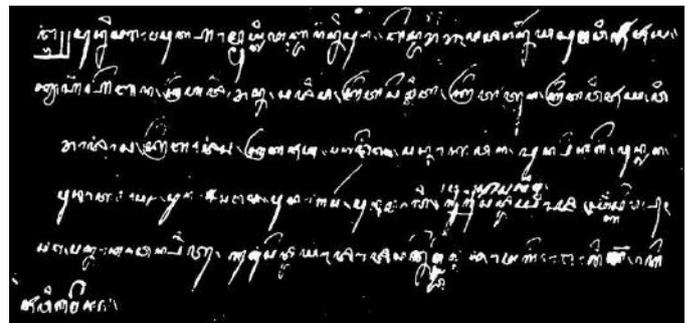
Gambar 10. Citra gray dari gambar uji coba 4



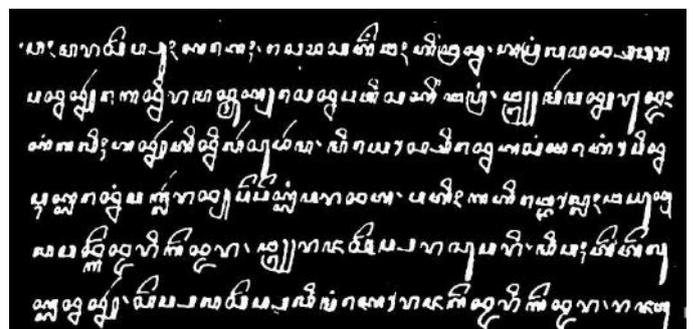
Gambar 13. Citra biner dari gambar uji coba 2



Gambar 11. Citra gray dari gambar uji coba 5



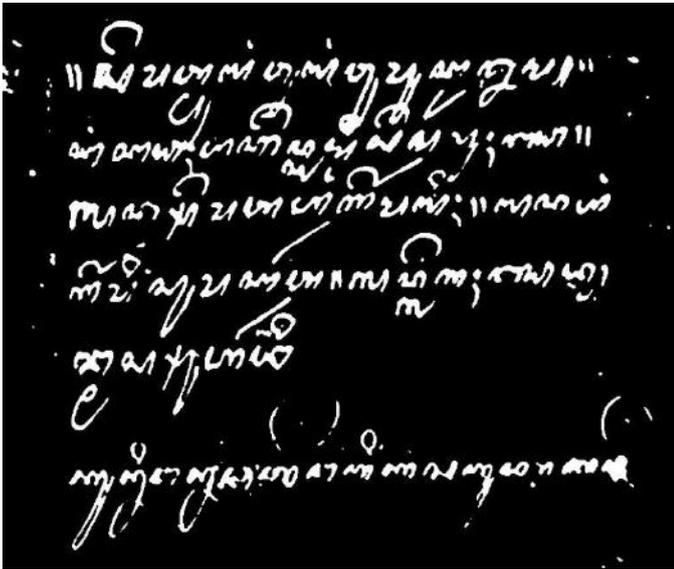
Gambar 14. Citra biner dari gambar uji coba 3



Gambar 15. Citra biner dari gambar uji coba 4

C. Metode Otsu

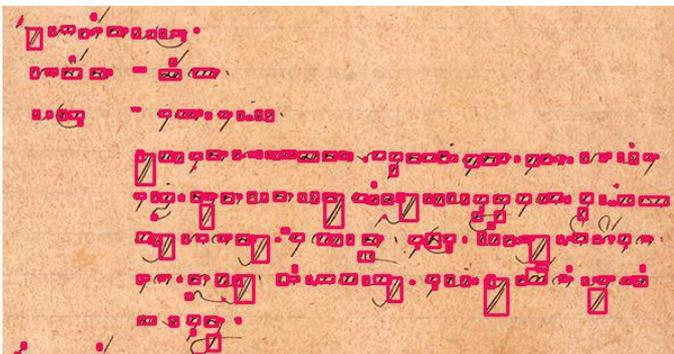
Dengan metode otsu, citra gray dapat diubah menjadi citra biner. Hasilnya adalah seperti berikut



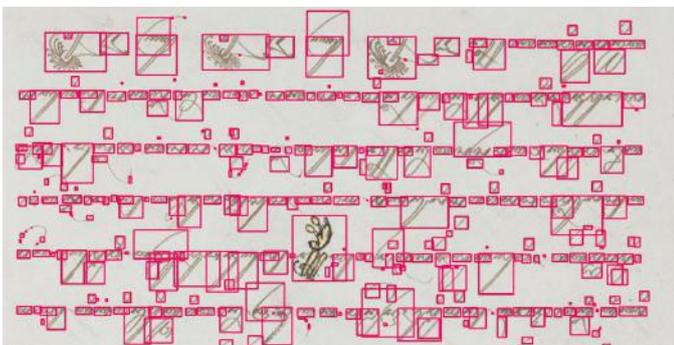
Gambar 16. Citra biner dari gambar uji coba 5

D. Pelabelan Aksara

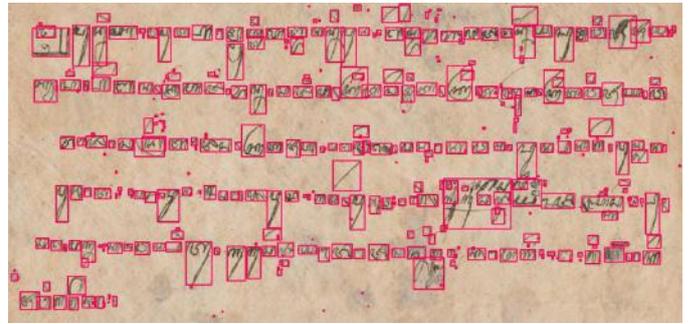
Setelah menjadi citra biner, maka selanjutnya adalah pelabelan. Hasilnya berupa penandaan pada citra yang dapat dilihat sebagai berikut.



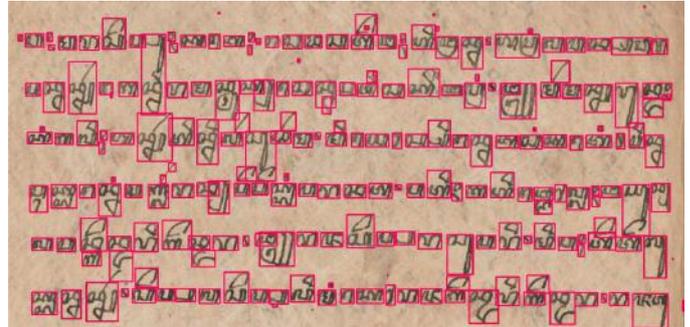
Gambar 17. Hasil pelabelan uji coba 1



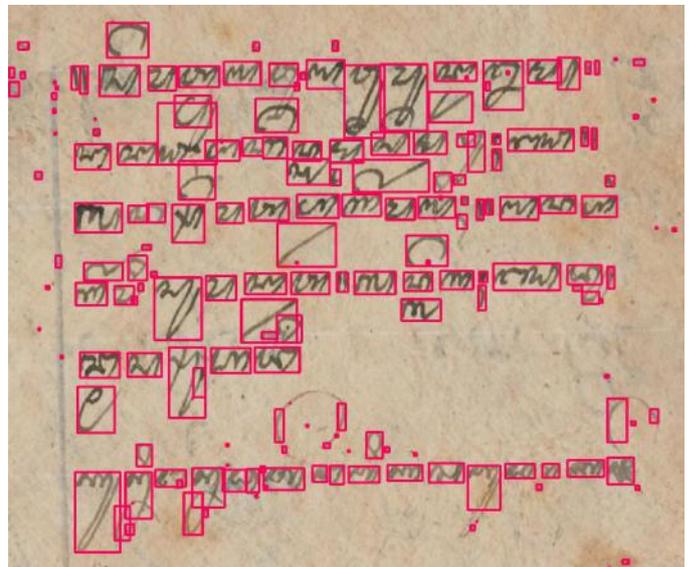
Gambar 18. Hasil pelabelan uji coba 2



Gambar 19. Hasil pelabelan uji coba 3



Gambar 20. Hasil pelabelan uji coba 4



Gambar 21. Hasil pelabelan uji coba 5

IV. KESIMPULAN

Dari hasil terlihat ada beberapa huruf yang sudah benar terdeteksi, namun ada beberapa objek yang bukan merupakan aksara tetapi tetap terlabelkan. Ini kemungkinan karena masih terdapat noise yang tidak hilang pada saat *preprocessing*. Deteksi Aksara Jawa dengan menggunakan *Otsu's Binerization* dapat menghasilkan hasil yang memuaskan jika

saat proses *preprocessing* menghilangkan *noise* dengan benar dan bersih.

Untuk selanjutnya mungkin akan dilakukan:
Menggunakan teknik *preprocessing* yang lebih bagus serta akan dilanjutkan dan dikembangkan menjadi aplikasi identifikasi aksara jawa sehingga Aksara Jawanya dapat ditranslasi menjadi huruf alfabet atau latin.

REFERENCES

- [1] Saraswati, Ufi. (2011, Maret). Arti dan Fungsi Naskah Kuno Bagi Pengembangan Budaya dan Karakter Bangsa melalui Pengajaran Sejarah. Dipresentasikan pada Seminar Nasional Pendidikan Sejarah, APPS di Bandung.
- [2] Putra, I. D. (2004). Binerisasi citra tangan dengan metode otsu. Majalah Ilmiah Teknologi Elektro, 3(2)