

Segmentasi karakter pada citra manuskrip Jawa menggunakan projection analysis dan connected component analysis = Character segmentation in javanese manuscript based on projection analysis and connected component analysis

Muhammad Zidny Naf'an, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20404136&lokasi=lokal>

Abstrak

[ABSTRAK

Penelitian mengenai segmentasi karakter merupakan proses awal dalam melakukan analisis isi naskah manuskrip Jawa. Segmentasi karakter pada citra naskah sangatlah penting untuk proses transliterasi naskah beraksara Jawa menjadi huruf Latin secara otomatis oleh komputer.

Kertas pada naskah sudah rapuh dikarenakan umur naskah yang tua sehingga menyebabkan kontras yang rendah dan munculnya noise antara teks dan background. Serta warna tinta yang sudah memudar berpengaruh pada kontras karakter pada naskah. Dimungkinkan juga terjadi ambigu hasil segmentasi karakter dikarenakan adanya keanekaragaman ukuran karakter dan adanya jarak spasi pada satu karakter.

Langkah-langkah segmentasi karakter yang dilakukan dalam penelitian ini adalah image acquisition, Gaussian filtering, binerisasi citra dengan algoritma Wolf-Jolion, segmentasi baris dengan metode vertical profile projection (VPP) dan Radon transform serta rule deteksi objek dengan connected component analysis, kemudian segmentasi karakter dengan metode horizontal profile projection.

Evaluasi segmentasi baris dilakukan dengan perhitungan nilai hit rate untuk empat kriteria, yaitu Correct untuk hasil yang benar, Line Including Outlier Script (LIOS) untuk hasil segmentasi yang memiliki aksara dari baris lain, Truncated Line (TL) untuk baris yang terpotong, dan Mix (LIOS+TL) untuk baris yang memiliki aksara dari baris lain dan juga terpotong. Hasil segmentasi baris yang benar (Correct) dengan metode VPP adalah 0.77 dan dengan metode Radon transform adalah 0.8. Hal ini dikarenakan perhitungan jumlah piksel hitam pada Radon transform berdasarkan sudut tertentu, sedangkan pada VPP perhitungan jumlah piksel hitam hanya berdasarkan satu baris saja.

Evaluasi hasil segmentasi karakter dilakukan dengan perhitungan nilai hit rate untuk dua kriteria, yaitu Readable untuk karakter yang dapat dibaca dan Unreadable untuk karakter yang tidak terbaca. Hasil segmentasi karakter dengan citra baris hasil dari vertical profile projection adalah 0.60245 dan nilai hit rate hasil segmentasi karakter dengan citra baris dari Radon transform adalah 0.60036. Hasil segmentasi karakter dengan input citra baris hasil segmentasi Radon transform hasilnya lebih rendah dikarenakan terdapat beberapa hasil segmentasi baris yang tumpang tindih (2 baris tersegmentasi menjadi 1 baris). Sehingga horizontal profile projection gagal dalam melakukan identifikasi lokasi karakter pada baris tersebut.;

ABSTRACT

Research on the segmentation of character is the initial process in conducting an analysis of the Javanese

manuscript. Character segmentation in the image of a manuscript is essential to the process of transliterating Javanese script into Latin characters automatically by computer.

Paper on a manuscript already fragile due to age old script causing the low contrast and the emergence of noise between text and background. As well as the color of the ink that's been fading effect on contrasting characters in the script. It is possible occur ambiguous character segmentation results due to the diversity of character size and distance of space on a single character.

Segmentation character steps that are undertaken in this research are image acquisition, Gaussian filtering, image binarization with Wolf-Jolion algorithm, segmentation line with a method of vertical profile projection (VPP) and Radon transform and the rule detection of objects by connected component analysis, then segmentation characters by horizontal profile projection method.

The evaluation of the segmentation of the line carried out by calculation the hit rate to four criteria , that is Correct for result right, Line Including Outlier Script (LIOS) for the results of the segmentation of having script of the other line, Truncated Line (TL) for line truncated, and Mix (LIOS + TL) for line having script of the other line and also cut off. The results of segmentation line are properly (Correct) by VPP method is 0.77 and with Radon transform method is 0.8. It was because the calculations the number of pixels black on radon transform based on certain angles while in VPP calculating the number of black pixels only on one line.

Evaluation of the results of the segmentation of the character performed with calculation of the value of a hit rate to two criteria namely Readable for characters that can be read and Unreadable for the character of being illegible. The results of the segmentation of character with the line images result of the vertical profile projection is 0.60245 and hit rate value as results of the segmentation of character with the line images result of the Radon transform is 0.60036. The results of the segmentation of character with the line images result of the Radon transform more low because there are the outcome of several segmentation lines that overlap (2 lines segmented become a single line). So that horizontal profile projection fail in doing identifying location of the characters on the line.

, Research on the segmentation of character is the initial process in conducting an analysis of the Javanese manuscript. Character segmentation in the image of a manuscript is essential to the process of transliterating Javanese script into Latin characters automatically by computer.

Paper on a manuscript already fragile due to age old script causing the low contrast and the emergence of noise between text and background. As well as the color of the ink that's been fading effect on contrasting characters in the script. It is possible occur ambiguous character segmentation results due to the diversity of character size and distance of space on a single character.

Segmentation character steps that are undertaken in this research are image acquisition, Gaussian filtering, image binarization with Wolf-Jolion algorithm, segmentation line with a method of vertical profile projection (VPP) and Radon transform and the rule detection of objects by connected component analysis, then segmentation characters by horizontal profile projection method.

The evaluation of the segmentation of the line carried out by calculation the hit rate to four criteria , that is Correct for result right, Line Including Outlier Script (LIOS) for the results of the segmentation of having script of the other line, Truncated Line (TL) for line truncated, and Mix (LIOS + TL) for line having script of the other line and also cut off. The results of segmentation line are properly (Correct) by VPP method is 0.77 and with Radon transform method is 0.8. It was because the calculations the number of pixels black on radon transform based on certain angles while in VPP calculating the number of black pixels only on one line.

Evaluation of the results of the segmentation of the character performed with calculation of the value of a hit rate to two criteria namely Readable for characters that can be read and Unreadable for the character of being illegible. The results of the segmentation of character with the line images result of the vertical profile projection is 0.60245 and hit rate value as results of the segmentation of character with the line images result of the Radon transform is 0.60036. The results of the segmentation of character with the line images result of the Radon transform more low because there are the outcome of several segmentation lines that overlap (2 lines segmented become a single line). So that horizontal profile projection fail in doing identifying location of the characters on the line.

]