

ABSTRAK

Perbandingan Akurasi Deteksi Baris Min-Cost Flow Network dan Tesseract Pada Citra Digital Berdasarkan Perbedaan Metode Segmentasi

Oleh

Andi Malia Fadilah Bahari

1197050014

Pada saat ini, *Optical Character Recognition (OCR)* telah berkembang hingga memiliki fokus penelitian yang berbeda berdasarkan tipe citra yang dideteksi, seperti *Natural Scene Image* dan *Document Image*. Pemfokusan berdasarkan tipe citra dilakukan karena tiap tipe citra memiliki masalah tersendiri sehingga memerlukan alur efektif pada tiap tipe. Pada penelitian terdahulu, masih terdapat deteksi teks dengan metode tradisional yang memungkinkan pendeteksian pada kedua tipe citra selama tidak mengalami distorsi. Sistem metode tradisional yang digunakan sendiri berfokus pada dua pendekatan yaitu *Sliding Window* dan *Connected Component Analysis* dalam proses deteksi karakter kandidatnya dengan penggunaan metode segmentasi yang beragam. Dikarenakan sistem-sistem tersebut belum pernah dibuktikan pengujiannya terhadap dataset *Document Image* dengan kandungan teks yang relatif banyak. Salah satu penelitian metode tradisional dengan penggunaan *Min-Cost Flow Network (MCFN)* untuk merapihkan barisan karakter dan *Tesseract* sebagai filter akhir deteksi karakter kandidat terhadap *Document Image*. Penelitian *MCFN* tersebut memberikan hasil yang lebih rendah dibandingkan pengujian terhadap *Natural Scene Image* pada penelitian terdahulu, dikarenakan permasalahan pada proses segmentasi karakter. Penelitian ini berfokus pada hasil akurasi deteksi baris teks metode tradisional yang menggunakan *MCFN* dan *Tesseract* pada infografis dengan mengintegrasikan metode segmentasi yang berbeda-beda untuk mengetahui metode segmentasi yang dapat memberikan hasil akurasi terbaik. Dari penelitian ini menghasilkan rata-rata akurasi tertinggi didapatkan dari penggunaan metode segmentasi *Connected Component Labeling* dengan hasil *Recall* 35,29%, *Precision* 69,10%, dan *F-Score* 41,02% pada parameter T_S jarak antara karakter sebesar 2.

Kata kunci: *Optical Character Recognition*, Segmentasi, Deteksi Baris, *MCFN*, *Tesseract*