

Abstrak

Digitalisasi dokumen fisik ke bentuk digital dapat dilakukan dengan dua cara. Pertama, ke bentuk citra yang digitalisasinya mudah. Namun, sulit dilakukan ekstraksi informasinya. Kedua, ke bentuk teks yang mudah diekstraksi informasinya. Namun, diperlukan proses khusus untuk mengenali teks. *Optical Character Recognition* (OCR) merupakan bentuk digitalisasi berkas fisik menjadi teks digital. Metode yang populer digunakan dalam OCR adalah Jaringan Syaraf Tiruan (JST) dan *Hidden Markov Model* (HMM). Namun, JST membutuhkan waktu kerja yang lama dan jarang sekali ditemukan laporan tentang waktu kerja HMM.

Pada tugas akhir ini, digunakan metode *Fast Wavelets Transform* (FWT). Pengenalan karakter dilakukan dengan membandingkan *Decomposition Coefficient Vector* (DCV) menggunakan pendekatan *Hierarchical Classification* berdasarkan koefisien yang diikuti dalam klasifikasi. Dilakukan pula *error reduction* dengan perhitungan rasio untuk mengurangi misklasifikasi karakter yang bentuk kapital dan bentuk kecilnya sama.

Hasil pengujian menunjukkan bahwa FWT dengan model satu *typeface* menghasilkan akurasi yang tinggi di lingkup terkecil pengujian. Sedangkan untuk lingkup yang lebih besar, diperlukan kombinasi *typeface* untuk hasil yang lebih baik. Namun, pembangunan model dengan kombinasi *typeface* memakan waktu lebih lama dari pembangunan model satu *typeface*. Analisis hasil pengujian menunjukkan bahwa konsep *hierarchical classification* berhasil mereduksi ukuran DCV pada 90,278% karakter dan *error reduction* meningkatkan akurasi sistem sebanyak 2,712%.

Kata Kunci: *Optical Character Recognition, OCR, Fast Wavelet Transformation, Wavelet, Hierarchical Classification.*