ABSTRAKSI

Optical Character Recognition (OCR) adalah teknik yang digunakan untuk menerjemahkan suatu citra yang memuat karakter atau huruf menjadi dokumen teks dalam standar pengkodean seperti ASCII atau Unicode. Biasanya OCR mengambil masukan berupa citra hasil scan dari dokumen tercetak dan melakukan pengenalan pada masing-masing karakter dalam citra tersebut. Dengan demikian OCR mengubah citra masukan menjadi teks digital yang dapat diedit sebagaimana umumnya.

Dalam melakukan pengenalan karakter sistem OCR melakukan empat langkah utama yaitu pre-prosesing, segmentasi, ekstraksi fitur, dan pengenalan karakter. Pre-prosesing bertujaun untuk mempersiapkan citra masukan dengan cara memperbaiki kualitasnya atau mereduksi noise sehingga lebih mudah diinterpretasi pada tahap selanjutnya. Segmentasi adalah proses pemenggalan karakter-karakter dalam citra menjadi satuan-satuan lebih kecil berupa huruf. Segmen-segmen ini kemudian diekstrak cirinya melalui proses ekstraksi ciri dan dikenali sebagai karakter-karakter tertentu pada tahap pengenalan.

Diskusi-diskusi tentang OCR lebih banyak terfokus pada proses ekstraksi ciri dan pengenalan. Tugas akhir ini berusaha menerapkan proses OCR dengan menerapkan Principal Component Analysis (PCA) dan K-Nearest Neigbourhood (KNN) masing-masing sebagai metode ekstraksi ciri dan metode pengenalan. Oleh karena itulah penulis memilih judul *"Optical Character Recognition (OCR) menggunakan Principal Component Analysis (PCA) dan K-Nearest Neighbourhood (KNN)"*

Kata kunci: Optical Character Recognition (OCR), Principal Component Analysis (PCA), K-Nearest Neighbourhood (KNN)"