ABSTRACT

Optical Character Recognition (OCR) is a technic used to translate images that contain characters, letters or both, become text document within the standard of ASCII or Unicode. Commonly, OCR takes scanned images from printed documents as input and identifies each character in the image. In the other words, OCR converts inputted images to simple digital text than can be editedable normally.

There are four main steps on the process of recognizing characters by OCR : pre-processing, segmentation, feature-extraction, and character-identifying. Pre-processing means to prepare inputted images by improving the quality or both reducing noises therefore simplifying next steps interpretation. Segmentaton is a process to separat each character on the images, become a single letter. The segmentations results are extracted on the feature-extraction and recognition as specific character on the character-identifying.

The main discussion of OCR is mainly focused on the process of feature-extraction and character-recognition. This final project is made using Principal Component Analysis (PCA) and K-Nearest Neigbourhood (KNN) as a feature-extracton method and identifying method. Therefore the writer choosing title : *"Optical Character Recognition (OCR) using Principal Component Analysis (PCA) and K-Nearest Neighbourhood (KNN)"*.

Keywords : Optical Character Recognition (OCR), Principal Component Analysis (PCA), K-Nearest Neighbourhood (KNN).