

Abstract

Digitization of physical documents can be done in two ways. First, to image that can be done easily but the information extraction in the document is rather difficult. Second, to digital text that its information can be extracted easily. However, it needs additional process in text recognition. Optical Character Recognition (OCR) is a process that digitize physical document to digital text. Popular methods used in OCR are Artificial Neural Network (ANN) and Hidden Markov Model (HMM). However, ANN takes long time to run and reports about running time in HMM are rarely found.

This thesis applied Fast Wavelet Transform (FWT). The recognition is done by comparing Decomposition Coefficient Vector (DCV) using Hierarchical Classification approach based on what coefficient will be included in classification. Error reduction is also conducted with ratio comparison to minimize misclassification that can be happened between capital and normal characters that have same shape.

Results show that FWT using one typeface model produce high accuracy in the smallest scope of testing. As for the larger scope, it takes model that consist of combination of typefaces for better results. However, the model that use combination of typefaces takes longer time to build than the model with just one typeface. Analysis of the results show that Hierarchical Classification managed to reduce the size of DCV in 90,278% used character. Also, error reduction improves the accuracy of the system as much as 2,712%.

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