**ABSTRAK**

**PEMBELAJARAN HURUF DAN KALIMAT ARAB BERBASIS PENGENALAN CITRA**

**Oleh**

**Ainatul Radhiah**

**NIM: 23215145**

**PROGRAM STUDI MAGISTER TEKNIK ELEKTRO**

Optical Character Recognition (OCR) berfungsi untuk memindai citra digital menjadi teks yang ada didalam citra tersebut. Dengan adanya OCR, citra yang berisi tulisan tangan, tulisan mesin ketik atau computer text, dapat dimanipulasi. Bahasa arab digunakan oleh lebih dari 1 miliyar pengguna di dunia. Jika sistem OCR tersedia untuk karakter arab, akan memiliki manfaat yang banyak. Telah banyak penelitian tentang pengenalan karakter latin, Jepang dan Cina, tetapi sedikit penelitian tentang huruf Arab, ini dikarenakan terdapat kesulitan dalam huruf Arab, yakni pada bagian segmentasi, karakteristik tulisan Arab yang bersambung membuat tahap segmentasi huruf menjadi sulit. Pada penelitian ini akan dirancang sistem yang dapat mengenali huruf arab dalam rangkaian kata. Adapun tahap yang dilakukan ialah *pre-processing*, segmentasi, *thinning*, ekstraksi fitur, dan klasifikasi. Pada tahap *pre-processing* dimulai dengan *binarization, skew detection, baseline estimation, median filter* dan *resizing*. Pada tahap segmentasi, segmentasi baris dilakukan dengan verifikasi lintas *horizontal projection* dengan standar deviasi, segmentasi kata dilakukan dengan komponen pixel yang terhubung, dan segmentasi huruf dilakukan dengan algoritma Zidouri. Pada tahap *Thinning* dilakukan dengan algoritma Zhang Suen. Pada tahap ekstraksi fitur dilakukan dengan metode *Freman Chain Code,* jumlah komponen yang terhubung, jumlah lubang, jumlah titik dan posisi titik. Pada tahap klasifikasi menggunakan metode *Hidden Markov Model* (HMM), dan *Support Vector Machine* (SVM) untuk dibandingkan hasil dari kedua metode tersebut. Hasil dari semua pengujian menunjukkan bahwa klasifikasi dengan HMM menghasilkan akurasi sebesar 50% dan klasifikasi dengan SVM menghasilkan akurasi sebesar 85%.

Kata kunci: OCR, *Freeman Chain Code*, HMM, SVM

**ABSTRACT**

**LEARNING ARABIC LETTERS AND SENTENCE BASED ON IMAGE RECOGNITION**

**By**

**Ainatul Radhiah**

**NIM: 23215145**

**ELECTRICAL ENGINEERING MASTER PROGRAM**

Optical Character Recognition (OCR) allows you to scan a digital image into text. With OCR, the image of handwritten text, writing text typewriter or computer, can be manipulated. Arabic is used by more than 1 billion users worldwide. If the OCR system is available for Arabic characters, will have many benefits. It has been a lot of research on Latin character recognition, Japan and China, but a little research on the Arabic letter, is because there are difficulties in the Arabic letter, which is on the segmentation, the characteristics of the Arabic letters is connected making stage segmentation becomes difficult. This research will be designed a system that can recognize Arabic letters in a series of words. The stage is carried out pre-processing, segmentation, thinning, feature extraction and classification. In the pre-processing stage began with binarization, skew detection, estimation baseline, median filter and resizing. In segmentation stage, segmentation line is done with cross verification with a standard deviation of horizontal projection, segmentation of words is done by pixel components are connected, and segmentation letter is done by Zidouri algorithms. In Thinning stage is done by Zhang Suen algorithms. In the feature extraction stage was conducted by Freman Chain Code, the number of connected components, number of holes, the number of points and the position of dots. In the classification stage using Hidden Markov Model (HMM), and Support Vector Machine (SVM) to compare the results of the two methods. The results of all tests show that the classification by HMM produces an accuracy of 50% and a yield SVM classification with an accuracy of 85%.

Keywords: OCR, *Freeman Chain Code*, HMM, SVM.