

Dedication

To my great educator

Alsheikh Mohamed Mohana Basheir

to my lovely Mother, Father,
Sisters and Brothers.

ACKNOWLEDGEMENTS

Gratefully thanks to my supervisor Dr. Mohamed El Hafiz Mustafa for his supervision, valuable comments and support throughout research period.

My thanks extended to all SUST- ARG (Sudan University of Science and Technology- Arabic Recognition Group) members who collect the data needed for this research.

ABSTRACT

Hidden markov models (HMMs) are stochastic models which are widely used in speech recognition. Later HMMs began to be applied in handwriting recognition.

This thesis tested the performance of HMMs on Indian handwritten digits recognition using SUST-ARG-digits dataset. SUST-ARG digits dataset is a newly established Indian handwritten numeral dataset collected by Sudan University of Science and Technology Arabic Recognition Group. Features are extracted from digits images using chain coding method.

HMMs trained using Baum-Welch algorithm by a training dataset of size 1350 samples of the nine digits (150 samples for each digit). The best number of states has been searched experimentally. The experiments show that a model with 9 states gives the best results (97.78% recognition rate) on a testing data of size 180 of the nine digits (20 samples for each digit).

المستخلص

نماذج ماركوف الخفية هي عبارة عن نماذج عشوائية تستخدم بصورة واسعة في التعرف على الكلام وفي بعض التجارب الحديثة اظهرت هذه النماذج نتائج جيدة في التعرف على الكتابة.

في هذا البحث تم اختبار أداء نماذج ماركوف الخفية في التعرف على الأرقام العربية (الهندية) المكتوبة بخط اليد باستخدام مجموعة بيانات الأرقام المكتوبة بخط اليد (SUST- ARG) وهي عبارة عن مجموعة بيانات جمعت حديثا بواسطة مجموعة التعرف على الكتابة العربية بجامعة السودان للعلوم والتكنولوجيا .

تم استخلاص السمات من صور الأرقام باستخدام طريقة ترميز السلسلة.

تم تدريب نماذج ماركوف الخفية باستخدام خوارزمية Baum Welch بمجموعة بيانات حجمها 1350 عينة من تسعة أرقام (150 عينة لكل رقم).

اظهرت النتائج أن النماذج ذات التسع حالات هي التي تعطي أفضل النتائج لمجموعة البيانات المعطاة بمعدل تعرف 97.78 %.

Table of Contents

Title	Page
Dedication	I
Acknowledgements	II
Abstract	III
المستخلص	IV
Table of Contents	VI
List of Tables	VII
List of Figures	VIII
CHAPTER ONE	
1. INTRODUCTION	2
1.1 Research Problem	2
1.2 Objectives	2
1.3 Tools	3
1.4 Thesis Organization	3
CHAPTER TWO	
2. LITERATURE REVIEW	5
2.1 Pattern Recognition	5
2.1.1 Pattern Recognition Approaches	5
2.1.2 Pattern Recognition System Components	6
2.1.3 Pattern recognition system Development stages	7
2.2 Optical Character Recognition	9
2.2.1 Handwriting Recognition	9
2.2.1.1 Handwritten Digits Recognition	9
2.3 Most Significant Development in Arabic Handwritten digit Recognition	10

CHAPTER THREE	
3. HIDDEN MARKOV MODELS (HMMs)	13
3.1 Elements of Hidden Markov Model	14
3.2 HMMs Three Basic Problems	16
3.3 HMMs Basic Problems Solutions	17
3.3.1 Problem 1 Solution	17
3.3.2 Problem 2 Solution	18
3.3.3 Problem 3 Solution	19
3.4 HMMs Types	22
CHAPTER FOUR	
4. HMM DIGIT RECOGNIZER	24
4.1 SUST-ARG Digits Dataset	25
4.2 Preprocessing Step	26
4.3 Feature Extraction Step	28
4.4 HMM Digit Classifier	29
4.5 Experimental Results	32
CHAPTER FIVE	
5. CONCLUSION & FUTURE WORKS	30
5.1 Conclusion	36
5.2 Future works	37
References	39
Appendix	41

List of Tables

Title	Page
Table 4.1: The number of samples for each digit in (SUST-ARG) handwritten digits dataset	25
Table 4.2: The errors rates of testing digits for different number of states	33
Table 4.3: The confusion matrix when the number of states is 9	34

List of Figures

Title	Page
Figure 2.1: Pattern Recognition System Components	8
Figure 2.2: Arabic handwritten digits	10
Figure 3.1: Probabilistic parameters of a hidden Markov model	15
Figure 3.2: HMMs types	22
Figure 4.1: Arabic digits form	24
Figure 4.2: Illustration of digits segmentation and collection	25
Fig. 4.3: The preprocessing operations on digits	27
Figure 4.4: Examples of digit zero looks like digit one and five after preprocessing	28
Figure 4.5: The directions values in Freeman's chain code	28
Figure 4.6: HMM Parameters re-estimation based on Baum-Welch algorithm	30
Figure 4.7: Diagram for HMM digits classifier	31
Figure 4.8: The errors average rates with different number of states	32
Figure 4.9: Images of digit three recognized incorrectly by the HMM classifier	34