The People's Democratic Republic of Algeria
Ministry of Higher Education and Scientific Research
Unviversity Mohamed MOHAMED - M'SILA
Faculty of Mathematics and Informatics

COMPUTER SCIENCE DEPARTMENT

Case Number:.....

Research Paper

Presented in partial fulfillment of the Requirements for the Degree of Master

Field : Mathematics and Informatics

Branch : Computer Science

Specialty : Advanced Information Systems

By: Amroune Abdelheq

TOPIC

Postal Code Handwritten Recognition System

Publically defended on : 31 / 05 /2016 Before a Jury composed of:

B.Brahimi	**Université of M'sila**	**Chairman**
Dr. Assas Ouarda	**Université of M'sila**	**Supervisor1**
Dr. L Belabdelouahab-Fernini	**Université of M'sila**	**Supervisor2**
A.Therafi	**Université of M'sila**	**Examiner**

Class: 2015 /2016

Table of Contents

General Introduction

One of the most important effects the field of Cognitive Science that we have in the field of Computer Science is the development of technologies that make our tools more human. A very relevant present-day field of natural interface research is handwriting recognition technology. Handwriting number recognition is a challenging problem researchers had been investigating for so long especially in the recent years. In our study, there are many fields concerned with numbers, for example, checks in banks or numbers in car plates and letters, but our main focus was on the postal code for letters.

A system for recognizing isolated digits may be as an approach to deal with such an application. In other words, to let the computer understand the Arabic numbers that are written manually by users and views them according to the computer process. Scientists and engineers with interests in image processing and pattern recognition have developed various approaches to deal with handwriting number recognition problems such as, minimum distance, decision tree and statistics.

The main objective for our system is to locate, isolate and recognize Arabic digits that exist in letters. For example, different users had their own handwriting styles. Here the main challenge is let the computer system understand these different handwriting styles and recognize them as standard writing.

We present a system that deals with such a problem. The system starts by acquiring an image of some letters that contain digits. This image is digitized using an optical device. After applying some enhancements and modifications to the digits within the image, it can be recognized by using several algorithms.

The first chapter mainly deals with the optical recognition systems and the Arabic language and its unique characteristics. In addition, some of the problems that faced us during the preparation of this work are displayed.

Outline

In chapter two, we talk about letters and envelopes which are our main concern of image recognition and the Algerian postal system along side with some reviews of previous works on the same field of the image recognition process.

The third chapter gives an overview about the artificial neural network and the support vector machines, which are the methods we use to test our work.

The last chapter represents a full experiment of a the recognition process from acquiring the letter to the results obtained using Artificial Neural Network, Support Vector Machine and K-

Nearest Neighbor. Finally, this work end with a conclusion of the obtained results and the future perspectives.

General Conclusion

We developed a system for Arabic handwritten digits recognition. We efficiently chose a segmentation method to fit our demands. Our system successfully designs and implement a neural network which efficiently works without demands, the support machine vector implementation was success and did recognize digits, whereas K-nearest neighbor classifier had the highest recognition rate. In addition, the system is able to understand the Arabic numbers that were manually written by the user.

The achieved results demonstrates that the ANN based system has shown promising results despite the fact of being trained only on a single set of templates. number of nearest neighbor is increased when using the KNN classifier also the recognition rate is increase. The system has its advantages such as Less Time Complexity, Very Small Database and High- Adaptability to untrained inputs, with only a small number of features to calculate as compared to other methods, Yet, the system has a large scope for further developments, The system performance can be further increased by:

1) increasing the DATABASE used for training the ANN,SVM and KNN so as to enable it to recognize stylized fonts also.

2) using better algorithms for training the ANN, SVM and KNN so as to decrease the Time complexity while handling larger databases.

3) using a better Feature Extraction techniques so as to increase the precision of results.

4) increase the DATABASE used for wilaya recognition, so our system will be able to detect more than just the 48 wilaya of Algeria.

5) fuse the recognition techniques in order to get better results and increase the rate of recognition

References

[1] A. K. Jain, R. P. W. Duin and J. Mao, "Statistical Pattern Recognition: A Review", On Pattern Analysis and Machine intelligence, vol. 22, no.1, pp.4-37, Jan 2000.

[2] Amritpal kaur, Madhavi Arora , International Journal of Computer Science & Engineering Survey (IJCSES) Vol.4, No.5, pp 20-21,October 2013

[3] Atallah Mahmoud Al-Shatnawi and Khairuddin Omar, "Skew Detection and Correction Technique for Arabic Document Images Based on Centre of Gravity", Journal of Computer Science, vol. 5, pp. 363-368, 2009.

[4] B. M. Bushofa and M. Spann, "Segmentation and recognition of Arabic characters by structural classification", Image and Vision Computing, vol. 15, pp.167-179, 1997.

[5] Ching Y. Suen and Robert J. Shillman, "Low Error Rate Optical Character Recognition of

[6] Evelina Maria De Almeida Neves, Adilson Gonzaga, Annie France Frere Slaets, "A Multi-Font Character Recognition Based on its Fundamental Features by Artificial Neural Networks", IEEE, 1997.

[7] J. Mantas, "An Overview of Character Recognition Methodologies", Pattern Recognition, vol. 19, no. 6, pp. 425-430, 1986.

[8] H. Bunke, P. S. Wang and H. S. Baird (eds.), Hand Book of Character Recognition and Document Image Analysis, Singapore: World Scientific, 1994.

[9] Herbert F. Schantz, "The History of OCR: Optical Character Recognition", University of Michigan: Recognition Technologies Users Association, 1982.

[10] Khorsheed, M.S. Off-line Arabic character recognition-a review. Patt Anal.vol 5,pp 31-45. , 2002.

[11] L. Heutte, T. Paquet, J. V. Moreau, Y. Lecourtier and C. Oliver, "A Structural /Statistical Feature Based Vector for Handwritten Character Recognition", Pattern Recognition Letters, vol. 19, pp. 629-641, 1998.

[12] M. A. Mohamed and P. Gader, "Generalized Hidden Markov Models – Part II: Application to Handwritten Word Recognition", IEEE Trans. Fuzzy Systems, vol. 8, no. 1, pp. 82-95, 2000.

[13] M. Schuster and K. K. Paliwal, November 1997 , "Bidirectional recurrent neural networks", IEEE Transactions on Signal Processing, vol 45, pp 2673–2681.".

[14] Nazif Arica, " A system for the recognition of the printed Arabic characters", Master Thesis, Faculty of Engineering, Cairo University: Egypt, 1975.

[15]. O. D. Trier, A. K. Jain and T. Taxt,"Feature Extraction Methods for Character Recognition A Survey", Pattern Recognition, vol. 29, pp. 641-662. 1996.

[16] P. Foggia, C. Sansone, F. Tortorella and M. Vento, "Combining Statistical and Structural Approaches for Handwritten Character Description", Image and Vision Computing 17, Elsevier Science, pp.701-711, 1999.

[17] R. J. Ramteke, "Recognition of Handwritten and Typed Based Document in Marathi", Ph.D Thesis, BAMU University: India, 2006.

[19] Rafael Gonzalez, C., E. Richard Woods and L. Steven Eddins, Digital Image Processingusing MATLAB. 2nd Edn., Prentice Hall, USA, pp: 624, 2003.

[20]. Sang Sung park,Won Gyo Jung,Young Geun Shin,Dong Sik Jang, Department of Industrial system and Information Engineering, Korea University, South Korea, "Optical character system Using BP Algorithm, vol.8 No.12, December 2008.

[21] Wikipedia , https://en.wikipedia.org/wiki/Microsoft_Wordc, consulted , 27/09/2015

[22] Wikipedia , https://en.wikipedia.org/wiki/Microsoft_PowerPoint, consulted 27/09/2015

[23] wikipedea, https://en.wikipedia.org/wiki/MATLAB, consulted 27/09/2015

[24] : wikipedea https://en.wikipedia.org/wiki/Windows_8.1, consulted 27/09/2015

[25] wikipedea, https://en.wikipedia.org/wiki/Envelope, consulted 15/03/2016

[26] wikipedea, https://en.wikipedia.org/wiki/Letter_(message), consulted 15/03/2016

[27] wikipedea https://en.wikipedia.org/wiki/Alg%C3%A9rie_Poste, consulted 15/03/2016

[28] wikipedea : https://en.wikipedia.org/wiki/List_of_postal_codes_in_Algeria, consulted 15/03/2016

[29]. Yusuf Perwej, " Recurrent Neural Network Method in Arabic Words Recognition System" ,International Journal of Computer Science and Telecommunications (IJCST), published by Sysbase Solution (Ltd), UK, London (http://www.ijcst.org) , vol. 3, Issue 11, pp 43-48, 2012.

[30] Z.Huang, & Leng, J. Analysis of Huâ€™s Moment Invariants on Image Scaling and Rotation. Proceedings of 2010 2nd International Conference on Computer Engineering and Technology (ICCET). (pp. 476-480). 2010

ملخص – اعتمدنا في إنشاء هذا النظام على ثلاث مصنفات. الشبكة العصبية ذات طبقة واحدة خفية، مصنف شعاع الدعم الآلي ،و مصنف ن-أقرب جار و هذا قصد انشاء نظام التعرف على أرقام الرموز البريدية مستخدمين طريقة عزوم هو. تم تدريب النظام وتقييمه على عدة أشكال مختلفة من نماذج خط اليد المقدمة من طرف المشاركين من الذكور والإناث. اثبتت تجارب اختبار ان الحجم الرموز البريدية يؤثر على دقة التعرف، بالإضافة لتأثير أسلوب الكتابة اليدوية. وأظهرت النتائج أن أسلوب الكتابة اليدوية قد تتفاوت وتؤاثر على دقة التعرف التي تسلط الضوء على بعض المشاكل مع ترميز الأرقام البريدية.

الكلمات المفتاحية: الشبكة العصبية, شعاع الدعم الآلي, ن-أقرب جار, عزوم هو, الرموز البريدية, نظام التعرف.

Abstract — A three based classifiers system was created. A back-propagation neural network with one hidden layer , a support Vector machine classifier, and K- nearest Neighbor classifier were used to create an adaptive postal code digits recognition system by using the Hu moments invariants feature extraction method. The system was trained and evaluated through different forms of handwriting samples provided by both male and female participants. Experiments tested, the effect of the size set on the recognition accuracy, and the effect of handwriting style on the recognition accuracy. Results showed that the handwriting style of the subjects had varying and drastic effects on the recognition accuracy which allowed to identify some of the problems with the system digits encoding.

Keywords : KNN, SVM, ANN, Hu moments ,code postal, recognition system.

Résumé- Un système basé sur trois classificateurs a été créé. Un réseau neuronal back-propagation avec une seule couche cachée, et une machine à vecteurs de support classificateur, et des plus proches voisins de classificateurs k ont été utilisés pour créer un système adaptatif de reconnaissance de chiffres du code postal, en utilisant la méthode d'extraction des caractéristiques des moments de Hu. Le système a été formé et évalué à travers les différentes formes d'échantillons d'écriture fournies par les deux participants masculins et féminins. Les expériences ont testé, l'effet de la taille indiquée sur la précision de la reconnaissance, et l'effet du style d'écriture sur la précision de la reconnaissance. Les résultats ont montré que le style d'écriture des sujets avait des effets drastiques et variables sur la précision de la reconnaissance qui a permis d'identifier certains des problèmes avec l'encodage des chiffres du système.

Les mots clés : RNA, MVS, KNN , les moments de Hu, code postal, système de reconnaissance