ASEE AMERICAN SOCIETY FOR ENGINEERING EDUCATION

Journal of Engineering Technology
(ISSN: 0747-9964)
Volume 5, Issue 2, July, 2016, PP.156-168

ORIGINAL ARTICLE

Script Segmentation and Classification based on Neural Networks versus Heuristics Approaches

Tanzila Saba

College of Computer and Information Sciences

Prince Sultan University Riyadh 11586 Saudi Arabia

Abstract

Neural Networks are widely applied to document analysis and classification. Nonetheless, research for printed text classification is well matured, several other handwritten historic document processing tasks such as preprocessing, character segmentation, script classification, signature and writer identification/verification are still on the way to its maturation. In this regard, there are two schools of thoughts first favor Neural Network based approaches and second like heuristics rule based approaches for documents analysis and classification tasks. Nonetheless, both groups have their justifications and achievements. This paper surveys the most significant problems in the area of OCR with and without artificially trained tools such as Neural Networks, hidden Markov models The objective of this paper is to provide a comparative study between Neural Networks and heuristics techniques employed for OCR in the state of the art. A particular emphasis is the contemporary role of training for the operation. Finally, the paper provides a critical analysis with respect to OCR, segmentation accuracy, error rate, single or fused classifiers being employed on benchmark databases. Additionally, the most promising research guidelines in the current field are recommended.

Keywords: ANN training, script classification, heuristic approaches, supervised learning, non-supervised learning, standard databases.

1. Introduction

Despite the wide use of electronic communication, paper document such as data entry forms, postal envelopes, and checks have central importance in our daily lives. As paper documents are cheap, reliable, secure for future reference, easily available and flexible in filling. Consequently, paper documents produced presently are more than ever before (Saba et al., 2015). Additionally, most of the governments and private organizations use paper based documents to collect information. However, it is laborious, time consuming to collect handwritten information from forms and typed it into computers by human operator. Whereas, electronically processed documents are easy to process for searching, updating and for further processing. So digitizing the paper documents is worth. Accordingly, automation of this procedure has attracted intensive research work in this field (Rehman and Saba, 2014). Unfortunately, even after decades of intensive research efforts in this domain, capabilities of the current OCR systems are still quite limited for printed text only and a small fraction of the data are entered into the computers automatically. A complete solution to the automatic cursive script extraction and its classification from documents has not yet been investigated (Saba et al., 2016).

In process of document analysis and classification, character segmentation is of vital importance. Character segmentation as part of cursive handwriting classification system has surveyed in (Saba et al., 2014). Specific

ASEE AMERICAN SOCIETY FOR ENGINEERING EDUCATION
ORIGINAL ARTICLE
Journal of Engineering Technology
(ISSN: 0747-9964)
Volume 5, Issue 2, July, 2016, PP.156-168

surveys about character segmentation techniques for both printed and cursive characters have been reported in (Saba et al., 2015). However, recent published research on character segmentation of cursive handwriting has not yet surveyed in presence of artificial neutral networks/heuristics rule based approaches and size of train/test data. Hence, this paper compares how artificial Neural Network effect the performance of character segmentation processes as well as what are the effects of size of train/test database. Some of segmentation techniques in this survey are part of handwriting classification system and the others are as an independent character segmentation system (Rehman and Saba, 2013; Kurniawan et al., 2011; Fadhil et al., 2016).

The further paper is organized in three sections. Section 2 presents segmentation systems based on Neural Network and heuristics rule based approaches. These systems are described step by step and their performance is evaluated regarding the use of benchmark/non-benchmark database, training and testing data size. In section 3, detail comparisons are performed into two main categories that include database size and existence/non-existence of artificially Neural Network learning process. Finally, conclusion and future directions are set in the section 4.

2. Character Segmentation

Since last few decades many researchers have proposed novel character segmentation techniques. However, the first common step is to find prospective segmentation point (PSP) heuristically. Then to reduce over- segmentation, some of them include artificial Neural Network to identify correct/incorrect segmentation points and the others employ optimum path algorithm (Rehman et al., 2009a). This section describes several comprehensive approaches categorized into two parts: artificially learned and non-learned system. The main motivation is to analyze how learning and non-learning system affect the techniques' performance. Additionally, size of database for training and testing the techniques is also taken into account.

2.1 Neural Network based Approaches

Neural-based technique for segmenting cursive script is originated by (Eastwood et al., 1997). Whereas, a Neural Network is trained with feature vectors that characterize possible segmentation points and also including features that represented the lack of a segmentation point. In this regard, the feature vectors are manually extracted from data set in CEDAR benchmark database. They report that the accuracy of the network on a test set was 75.9%. However, this preliminary study is still far from the expected result. The heuristic segmenter should enhance to obtain better possible segmentation points. Furthermore, the features vector is quite simple so might misclassify the correct and incorrect segmentation points (Saba and Rehman, 2012).

Few researchers divide segmentation process into two component, simple heuristic segmenter algorithm and feed-forward Neural Network using back-propagation algorithm (Rehman and Dzulkifli, 2008). Heuristic segmentation algorithm seeks minima or arcs between letters, then "hole-seeking" is used to handle incorrect segmentation on letters like "a", "u" and "o". Last step is to check segmentation point position that closest or greater than average width of some words. Incorrect segmentation is reduced by using neural component. Experiments utilize approximately 150 handwritten words are experimented and accuracy rate up to 78% and 66% respectively for one and two writers are claimed (Rehman et al., 2009b; Neamah et al., 2014).

Journal of Engineering Technology
(ISSN: 0747-9964)
Volume 5, Issue 2, July, 2016, PP.156-168

Blumenstein (2008) enhanced existing neural-based segmentation technique proposed by Blumenstein and Verma (2001). The existing segmentation technique adopted a feature-based heuristic segmenter (FHS) from Blumenstein and Verma (1999) to over-segment the word image. Afterward, a neural confidence-based module is applied to assess the possible segmentation point by calculating a fused value from three neural confidence values. The confidence values consist of segmentation point validation (SPV), left character validation (LCV) and centre character validation (CCV). In order to obtaining better possible segmentation points, enhanced heuristic segmenter (EHS) was adopted. Moreover, the neural confidence- based module is enhanced by adopting robust feature extraction technique (Blumenstein et al., 2004) for the classifier of left and centre characters. Regarding to improve segmentation accuracy, segmentation path detection-based character extraction technique is also performed (Cheng et al., 2004; Saba et al., 2014). Blumenstein claimed that his technique could reduce number of missed segmentation points and further decrease the processing time of whole stages.

Moreover, fusion of several ANN might improve the segmentation accuracy (Blumenstein and Verma, 2001). In this regard, authors adapt heuristic feature-based segmenter (Blumenstein and Verma, 1999) in order to generate prospective segmentation point (PSP). Afterward, unwanted PSP are removed by calculating confidence values of segmentation candidates with respect to ANN fusion. Three ANN are fused, where first ANN provides confidence value for correct and incorrect segmentation. Meanwhile, the two others are presenting value of left character confidence (LCC) and center character confidence (CCC). High confidence value means a good candidate for a segmentation point. Confidence values of correct segmentation point (CSP) and incorrect segmentation point (ISP) are computed by ANN.

Experiment is conducted using CEDAR benchmark database, under "BD/Cities" directory. They report that segmentation error for over-segmentation, missed, bad and bad+correct anchorage points are respectively 7.47%, 2.04%, 11.64%, and 6.3%. Even the over-segmentation significantly decreases after employing ANN fusion, yet bad segmentation is still high due to bottleneck of feature-based segmenter of previous work.

Verma (2002) performs character segmentation starting from baseline detection and generate over-segmentation points, extracted left, right characters and finally evaluate, join confidence values. Baseline is found by significant changes in the horizontal histogram density until reached the bottom image. Possible Segmentation points (PSP) are calculated using heuristic algorithm based on upper and lower word contours, holes, upper and lower contour minima, vertical density histogram and confidence assignment (Blumenstein, and Verma, 2001). Character segmentation problems that include part of another character are solved by tracing connected black pixels boundary of first and second segmentation points as illustrated in Fig. 2. To evaluate confidence values, three Neural Networks are employed (Verma and Gader, 2000). 300 words of "BD/Cities" directory of CEDAR benchmark database are experimented and accuracy rate 84.87% is claimed.

Cheng et al. (2004) extended cursive handwritten word segmentation process from their previous work of Blumenstein and Verma (1999). They also adopt validation scheme for left character (LC) and center character (CC) proposed by Verma (2002). Additionally, modified direction feature (MDF) that are proposed by Blumenstein et al. (2004) are added in order to enhance neural validation accuracy. MDF are calculated from background to foreground pixel, in both direction vertical and horizontal, with two values named left transition and destination transition (LT and DT). Moreover, problem of overlapped characters is solved

AMERICAN SOCIETY FOR ENGINEERING EDUCATION
Journal of Engineering Technology
(ISSN: 0747-9964)
Volume 5, Issue 2, July, 2016, PP.156-168
ORIGINAL ARTICLE

using segmentation path direction (SPD). SPD is used as base-fit line to determine when foreground pixels become starting point. In this regard, starting point must on top of base-fit line. Character extraction path can be discovered, if exploring to the right side that beginning from starting point were able to reach the upper row of images. However, when right path failed then exploring to left side from starting point is also conducted. Samples for experiments are taken from CEDAR benchmark database. They claim that percentage of correct segmentation using SPD on 317 words is 95.27%. Regarding to their study, MDF could improve the classifier so it become more sensitive. However, MDF extraction is increasing the computational complexity due to tracing calculation.

Cheng and Blumenstein (2005a) enhance heuristic segmenter (EHS) to improve Feature-based Heuristic Segmentation (FHS) algorithm by adding ligature detection proposed in (Blumenstein and Verma, 2001). Actually, FHS failed to determine segmentation point on overlapped characters. To determine ligature, EHS first calculate upper and lower baseline position to find main body. Then representation of modified vertical histogram counted distance between first black and last pixel in main body. Following histogram normalization, ligatures are determined by lowest value. EHS is examined using CEDAR database with 317 words. An accuracy rate 80.41% is reported.

Veloso et al., (2000) propose hypothesized character segmentation based on natural segmentation point and ligature detection. Natural segmentation referred to such character that are not connected and determined using histogram projection taken from five different angles. Ligature candidate obtained from morphological operations of opening and closing. While to search for the best structuring elements, genetic algorithm is applied to determine ligature in the set of training words. Finally Viterbi's algorithm used to determine qualified ligature. However, no accuracy rates are reported (Rehman and Saba, 2014)

Cheng and Blumenstein (2005a) improve previous work proposed by Cheng and Blumenstein (2005b) to determine EHS with segmentation based on Neural Network and PSP validation to improve character segmentation of cursive script. SPC, LCV and CCV are used for the validation of segmentation points. To get PSP precisely, EHS use ligature detection and neural assistance. Ligatures are determined by minimum value of modified vertical histogram in baseline. Neural assistant is sought to enhance accuracy. CEDAR benchmark database is used for training and experiment step. Error rate using EHS with neural assistant are such as over-segmented 7.37%, missed 0.1 %, and bad errors 6.79 %.

On the other hand, multiple validation modules after over-segmentation process are proposed by Lee and Verma (2008). Regarding to over-segmentation, vertical pixel density in the core-zone that less than threshold is assigned as over-segmentation points. Afterward, multiple validation module that based on holes detection, total foreground pixel between two neighboring segmentation points and Neural Network voting is introduced. In addition, missing segmentation point between neighboring characters is also checked. The experiments have been conducted on CEDAR benchmark database. Since slope angle correction was not performed, thus core-zone might mistakenly estimated then it might leads generated incorrect segmentation. Moreover, under segmentation error on their method was contributed by touching character that forms a holes-region, which is marked as incorrect segmentation points by holes detection algorithm.

Recently, Saba et al. (2010a) present a simple and effective character segmentation approach. Following preprocessing, characters geometric features are explored for over-segment characters. However, to identify

Journal of Engineering Technology
(ISSN: 0747-9964)
Volume 5, Issue 2, July, 2016, PP.156-168

incorrect segmentation, neural assistance is sought. Experiments are conducted on IAM benchmark database and character segmentation accuracy rate up to 91.21% is reported.

Neural Networks have been shown its maturity to validate segmentation points. Numerous researchers have been proven the effectiveness of validation scheme. According to recent studies, author considered to adopt such method in order to achieve encouraging segmentation accuracy. However, neural assistant seems to be time consuming to collect the training data and to train the ANN for the purpose of validation. The error rate only improved by 1% after applying segmentation and character validation using ANN fusion (Saba et al., 2010d). In addition, neural assistant utilize character recognizer to validate segmentation points that overburden the ANN with several input vector and demand high computation during validation. In this regard, neural validation is preferred and heuristic segmenter with robust rules and features should be investigated (Saba et al., 2010c).

2.2 Heuristic rule based approaches

Actually, there are several features in cursive scripts that are unaffected by handwriting variability, for example loops/holes, ligatures, ascenders and descenders. Some of researchers used these features to determine PSP(possible segmentation points) using heuristic algorithm (Saba et al., 2016). In this regards, several approaches exhibit dissimilar results (Rehman et al., 2014). This section served briefly general steps of each algorithm as well as its reported result.

Han and Sethi (1995) introduce heuristic segmenter based on geometrical features to classify letters in two groups, regular and singular. Segmentation scheme divided character into six classes. Letters boundaries identification based on several global coefficient, reference lines and words zone. Segmentation process started with two branch path, first path seek reference line and zone, to calculate ENC and EWC. Second path started by binary image to extract singular and regular features. Input image in second path is preprocessed into binary and skeleton image. Segmentation system tested using cursive handwritten of postal address consisted of 1119 words and reported accurate character segmentation rate is 85.7%.

Nicchiotti and Scagliola (2000) propose simple and effective segmentation methods. Initially, images are preprocessed to fix noise, slant and skew (Saba et al., 2015). Segmentation algorithm has three steps: detect possible segmentation points, stroke generation and stroke analysis. Possible segmentation points detection based on contour minima and holes features. On first step, ligature determined by finding minimum value of upper, medium and lower contours. Holes detection prevented segmentation on hole and could add segmentation points on its both sides. Closest PSP merged into a single PSP. Second step, seek best path to cut images into strokes, if cutting horizontally hit black pixel over a certain length, cut direction changed in the range $\pm 45°$ around vertical one. Finally strokes from step two checked for consistency, inconsistence strokes are discarded or merged with adjacent consistence one. To test algorithm, 850 words are used in the TRAIN directory of CEDAR database. Accurate character segmentation up to 86.9% is observed and ligature detection accuracy is 97.9%.

ASEE AMERICAN SOCIETY FOR ENGINEERING EDUCATION
ORIGINAL ARTICLE
Journal of Engineering Technology
(ISSN: 0747-9964)
Volume 5, Issue 2, July, 2016, PP.156-168

Yanikoglu and Sandon (1998) propose segmentation determined by evaluating cost function at each point along baseline using linear programming. First, single character extracted based on connected pixel using a region growing algorithm. Then, each point along baseline evaluated with cost function, linear programming used to determine weight. First lowest cost point is chosen as segmentation point. Using this point, separator line with certain angle cut the images. The angle is taken from 30, -20, -10, 0, 10, 20, and 30 based on dominant slant. Experiment performed on 750 words is written by 10 writers. Segmentation performance up to 92% is claimed.

Verma (2003) propose novel contour code feature in conjunction with rule-based segmentation based on contour characteristic like loop, hat shape, etc. Baseline computation is performed to determine upper, lower, middle, ascender and descenders line. Possible segmentation points (PSP) are detected by heuristic algorithm to observe changes on vertical histogram projection. Finally, correct segmentation is finalized by validating PSP with rules that detected loop, hat shape (‘^’ or ‘v’), remove two closest PSP and add segmentation point between two PSP that cross threshold value. Rule-based segmentation approach is tested on CEDAR database with test sample size of 1200 words. Errors rate is reported according to standard criteria that include over-segmentation 10.02 %, missed segmentation 0.2 % and finally, bad-errors 8.7% observed.

Blumenstein and Verma (1999) employ heuristic approach to segment hand-printed and cursive handwritten words. Authors proposed a heuristic algorithm to locate hole location, minima detection and projection pixel density vertically. Characters' segmentation accuracy up to 76.52% is reported.

Recently Rehman et al. (2008b) propose heuristic approach for character segmentation of cursive handwritten words. Accordingly, several heuristics criteria are prescribed to locate character boundaries. To apply proposed heuristic mechanism of character segmentation, characters are divided into three broad categories. The first category includes characters with up/down cusp of characters' shape such as u, v, w, m, n and second category include characters composed of holes/semi holes such as a, b, c, d, x etc. Finally, third category of characters is without holes/cusp such as t, l, r etc. Authors claimed performed experiments on IAM benchmark database and report high accuracy, speed as no training required.

3. Performance comparison

This section performs comparison among various character segmentation techniques reported in the state of art based on database used; sample training/testing size, preprocessing involvement, existence / non-existence of artificially trained tools and segmentation accuracy.

3.1 Database size

Performance of systems heavily depends on the size of database. Some segmentation algorithms have good results on small or medium database, but not encouraging results for large database. Table 1 exhibits performance comparison based on various factors. Related information consists of domain/database (DB), number of words (Sample), preprocessing (Pre.), and techniques employed. Number of trained/tested words is also an important criteria to evaluate the techniques and therefore split into small, medium and large numbers respectively with tens of words, with hundred of words and with thousands of words as criteria mentioned in (Saba et al., 2015).

ASEE AMERICAN SOCIETY FOR ENGINEERING EDUCATION

Journal of Engineering Technology
(ISSN: 0747-9964)
Volume 5, Issue 2, July, 2016, PP.156-168

ORIGINAL ARTICLE

Table 1. Summary of recent character segmentation methods

Reference	DB	Sample	Pre.	Methods	Features	SR
Small Database						
Verma *et al.* (1998)	Griffith University	50	t	FB, ANN	Lower contour, holes, character width	78.66
Blumenstein and Verma (1998b)	Griffith University	44	-	FB, ANN	Lower contour, holes, character width	63.64
Blumenstein and Verma (1998a)	CEDAR	33	t	FB, ANN	Lower contour, holes, character width	76
Sas and Markowska-Kaczmar (2007)	Polish	93	-	EA	grapheme	93
Medium Database						
Han and Sethi (1995)	Postal	50 env.	t, sl, th	FB, GB	Geometric, topologic features	85.7
Eastwood and Jennings (1997)	CEDAR	317	t, sl, th	ANN	Vertical histogram, holes, crossing histogram, upper contour, lower contour	75.9
Yanikoglu and Sandon (1998)	English	750	-	FB, LP	Connected component	92
Blumenstein and Verma (1999b)	CEDAR	724	t, sl	FB, 3CV, ANN	Holes, vertical histogram, character width, lower contour	76.52
Blumenstein and Verma (1999a)	CEDAR	317	t, sl	FB, ANN	Holes, lower contour, vertical histogram	81.21
Nicchiotti *et al.* (2000)	CEDAR	850	t, sl, sk	FB	Holes, lower contour, area, bounding box, stroke	86.9
Xiao and Leedham (2000)	CEDAR	200	-	KBE	Connected component, character structure	82.9
Kavallieratou *et al.* (2000)	English & Greek	500	t, sl, sk	TBL	Character width, vertical histogram	82
Blumenstein and Verma (2001)	CEDAR	300	-	FB, 3ANN	Holes, vertical histogram, character width, lower contour	78.85
Verma (2002)	CEDAR	300	-	FB, 3ANN, CE	Holes, vertical histogram, character width, lower contour	84.87
Maragoudakis *et al.* (2003)	English & Greek	500	sl, sk	FB, BBN	Character width, vertical histogram	86.4
Maragoudakis *et al.* (2003)	English & Greek	500	sl, sk	FB, NB	Character width, vertical histogram	83.8
Verma (2003)	CEDAR	1200 (char)	-	FB, RB	Holes, hat shape, character width, segmentation width	81.08
Cheng *et al.* (2004)	CEDAR	317	-	FB, 3ANN, MDF, CE	Holes, vertical histogram, character width, lower contour	95.27
Cheng and Blumenstein (2005a)	CEDAR	317	-	EHS, ANN, SPD	Holes, character width, lower contour, modified vertical histogram	82.54
Cheng and Blumenstein (2005b)	CEDAR	317	-	EHS, NA, MDF, SPD	Holes, character width, lower contour, modified vertical histogram	85.74
Lakshmi *et al.* (2006)	CEDAR	385	sl, th	HT	Hypergraph	-
Blumenstein (2008)	CEDAR	317	-	EHS, 3ANN, NA, MDF, SPD	Holes, character width, lower contour, modified vertical histogram	85.74
Lee and Verma (2008)	CEDAR	311	t	FB, 3ANN	Stroke thickness, vertical histogram, holes, total foreground pixel	81.45
Large Database						
Romeo-Pakker *et al.* (1995)	Arabic + Latin	1,383	-	CFA	Stroke thickness	93.5
Romeo-Pakker *et al.* (1995)	Arabic + Latin	1,383	-	FB	Upper contour	99.3

ASEE AMERICAN SOCIETY FOR ENGINEERING EDUCATION
Journal of Engineering Technology
(ISSN: 0747-9964)
Volume 5, Issue 2, July, 2016, PP.156-168
ORIGINAL ARTICLE

Preprocessing could be thresholding (t), slant (sl), skew (sk) correction or thinning/skeleton (th) for the methods briefly described as feature-based (FB), enhanced feature-based (EHB), rule-based (RB), geometric-based (GB), artificial Neural Network (ANN), neural assistant (NA), genetic algorithm (GA), Viterbi's algorithm (VA), character extraction (CE), Linear Programming (LP), Contour-following Algorithm (CFA), Bayesian Belief Networks (BBN), Naïve Bayes (NB). Noted for (RR) means this value is classification rates.

Some papers do not report number of experimented words. Therefore, they are reported in uncategorized part. Information on preprocessing step is also missing, so it is hard to know whether the sample is slanted or skewed. This information is important because for difficult words segmentation may lead many problems that may decrease segmentation accuracy. Some researcher use slant separator to cut the slanted word images (Rehman et al., 2012).

3.2 Neural Network vs. heuristic rule based approaches

Due to invariability of cursive script, few researchers implement segmentation validation using Neural Network or employ neural assistant to add segmentation points. On the other hand, feature based approaches are investigated over segmentation of cursive handwritten images. The character segmentation accuracy rates of these two techniques are quite encouraging. Table 2 presents performance comparisons of this survey.

Table 2. Neural Networks vs. heuristics approaches

Neural Network based approaches		
Reference	**ANN versus Heuristic systems**	**Accuracy (%)**
Blumenstein and Verma, (1997)	ANN	78.66
Blumenstein and Verma (1999)	ANN	76
Bozinovic and Srihari(1989)	ANN	81.21
Verma and Gader (2000	ANN	76.52
Blumenstein and Verma (2001)	ANN	78.85
Verma (2002)	Multiple A.I	84.87
Cheng, et al., (2004)	Multiple A.I	95.27
Cheng and Blumenstein (2005)	ANN + Heuristic	84.19
Cheng and Blumenstein (2005)	ANN + Heuristic	85.74
Lee and Verma (2008)	Multiple ANN	83.46
Rehman et al., (2012)	ANN	91.06
Rehman and Saba (2014)	ANN	88.08
Heuristic rule based approaches		

ASEE AMERICAN SOCIETY FOR ENGINEERING EDUCATION
ORIGINAL ARTICLE
Journal of Engineering Technology
(ISSN: 0747-9964)
Volume 5, Issue 2, July, 2016, PP.156-168

Reference	Methods	Accuracy (%)
Han and Sethi (1995).	Heuristic	85.7
Nishida and Mori(1992)	Heuristic	81.08
Nicchiotti and Scagliola (2000)	Heuristic	86.9
Verma (2003)	ANN + Heuristic	92
Yanikoglu and Sandon, (1998)	Linear programming	97
Maragoudakis et al.,2003	ANN + Heuristic	93.5
Rehman et al., (2008b)	Heuristics	89.62
Saba et al., (2015)	Heuristics	97.36

Table 2 shows a comparison and exhibits that segmentation rates of non-learning system are greater than learning system but this is not evidence that non-learning system is better than learning system. This graph just show average segmentation rates taken from several researches. Comprehensive experiment is needed to know how learning system affect segmentation rates. High segmentation rates are claimed by Saba et al., (2015) based on heuristic rules. However, size of data is an issue to be addressed in heuristic based approaches, indeed their accuracy is better than artificially trained tools but at the same time heuristic approaches are working on small data set. More experiments are needed to prove their worth in comparison to trained tools.

4. Conclusion and future directions

Neural Networks have been extensively used for document segmentation and classification in several systems. However, several issues rose to the efficiency of Neural Networks regarding training data, training time and slow speed of fused Neural Networks. Moreover, character segmentation using Neural Networks intelligence; validate characters candidates obtained from over-segmentation step by calculating confidence values. These techniques validate prospective segmentation as correct or incorrect segmentation points. Although, few researchers have employed more than one learning algorithms to achieve satisfactory results, however, there is still need to investigate these results in large databases. On one hand, trained system in large amount of data could lead the problem and are also time consuming. Although heuristic rules are used in segmentation process to over-segment the words and proved successful to locate the correct character boundaries, yet neural assistant may reduce further incorrect segmentation. In this regards, recursive Neural Networks may presents some of the most promising methods to be taken into consideration in future applications.

ASEE AMERICAN SOCIETY FOR ENGINEERING EDUCATION
Journal of Engineering Technology
(ISSN: 0747-9964)
Volume 5, Issue 2, July, 2016, PP.156-168
ORIGINAL ARTICLE

References

Blumenstein, M. (2008). Cursive Character Segmentation Using Neural Network Techniques. Machine Learning in Document Analysis and Classification, vol. 90, pp. 259-275.

Blumenstein, M., Liu, X. Y., and Verma, B. (2004). A Modified Direction Feature for Cursive Character Classification, Proceedings of the International Joint Conference on Neural Networks (IJCNN '04), Budapest, Hungary.

Blumenstein, M., Verma, B. (1999). A new segmentation algorithm for handwritten word classification. International Joint Conference on Neural Networks, (IJCNN '99), vol.4, pp.2893-2898.

Blumenstein, M., Verma, B. (2001). Analysis of segmentation performance on the CEDAR benchmark database. Proceedings of Sixth International Conference on document analysis and classification, pp.1142-1146. doi. 10.1109/ICDAR.2001.953964

Bozinovic, R. M., and Srihari, S. N. (1989). Off-Line Cursive Script Word Classification, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 11(8), pp. 68-83.

Cheng, C. K., and Blumenstein, M. (2005a). Improving the Segmentation of Cursive Handwritten Words using Ligature Detection and Neural Validation, Proceedings of the 4th Asia Pacific International Symposium on Information Technology (APIS 2005), Gold Coast, Australia, pp. 56-59.

Cheng, C. K., Liu, X. Y., Blumenstein, M., and Muthukkumarasamy, V. (2004). Enhancing Neural Confidence-Based Segmentation for Cursive Handwriting Classification, 5th International Conference on Simulated Evolution and Learning (SEAL '04), Busan, Korea, SWA-8.

Cheng, C.K., Blumenstein, M. (2005b). The neural-based segmentation of cursive words using enhanced heuristics. Proceedings of 8th International Conference on Document Analysis and Classification, vol. 2, pp. 650-654.

Eastwood, B., Jennings, A., and Harvey, A. (1997). A Feature Based Neural Network Segmenter for Handwritten Words. Proceedings of the International Conference on Computational Intelligence and Multimedia Applications. Gold Coast, Australia: 286–290.

Fadhil, MS. Alkawaz, MH., Rehman, A., Saba, T. (2016) Writers identification based on multiple windows features mining, 3D Research, vol. 7 (1), pp. 1-6, doi.10.1007/s13319-016-0087-6.

Han, K., and Sethi, I. K. (1995). Off-line cursive handwriting segmentation. Proceedings of third international Conference on Document Analysis and Classification, vol. 2, pp. 234-240.

Haron, H. Rehman, A. Adi, DIS, Lim, S.P. and Saba, T.(2012). Parameterization method on B-Spline curve. mathematical problems in engineering, vol. 2012, doi:10.1155/2012/640472.

Kavallieratou, E., Stamatatos, E., Fakotakis, N., and Kokkinakis, G. (2000). Handwritten character segmentation using transformation-based learning. Proceedings. 15th International Conference on Pattern Classification. 634 - 637.

ASEE AMERICAN SOCIETY FOR ENGINEERING EDUCATION
ORIGINAL ARTICLE
Journal of Engineering Technology
(ISSN: 0747-9964)
Volume 5, Issue 2, July, 2016, PP.156-168

Kurniawan, F. Rahim, MSM, Daman, D. Rehman, A. Dzulkifli, M. and Mariyam, S. (2011). “Region-based touched character segmentation in handwritten words”. International Journal of Innovative Computing, Information and Control vol. 7(6), pp.3107-3120.

Lakshmi, P. S., Hanmandlu, M., and Swaroop, A. (2006). Segmentation of Cursive Handwritten Words using Hypergraph. IEEE Region 10 Conference TENCON 2006. 1-4.

Lee, H., Verma, B. (2008) Over-Segmentation and Validation Strategy for Offline Cursive Handwriting Classification Proceedings of the International conference on intelligent servers, sensor Networks and Information processing, pp. 91-96

Maragoudakis, M. Kavallieratou, E. Fakotakis, N. and Kokkinakis, G. (2003). An Effective Stochastic Estimation of Handwritten Character Segmentation Bounds, ISAP 2003: Competitive Environment, Renewable Energy, Distributed Generation. Lemnos, Greece.

Nicchiotti, G., and Scagliola, C. (2000). A Simple and Effective Cursive Word Segmentation Method. Proceedings of the 7th International Workshop on Frontiers in Handwriting Classification, September, Amsterdam, ISBN 90-76942-01-3, Nijmegen: International Unipen Foundation, pp. 499-504.

Neamah, K. Mohamad, D. Saba, T. Rehman, A. (2014). Discriminative features mining for offline handwritten signature verification, 3D Research vol. 5(3), doi. 10.1007/s13319-013-0002-3.

Rehman, A. and Saba, T. (2014). Evaluation of artificial intelligent techniques to secure information in enterprises, Artificial Intelligence Review, vol. 42(4), pp. 1029-1044, doi. 10.1007/s10462-012-9372-9.

Rehman A., Kurniawan, F. and Mohammad, D.(2009a). Implicit Vs Explicit based Script Segmentation and Classification: A Performance Comparison on Benchmark Database, International Journal on Open Problems in Computer Science and Mathematics, vol. 2(3), pp. 352-364.

Rehman, A. Mohammad, D. Sulong, G. and Saba, T. (2009b). Simple and Effective Techniques for Core Zone Detection and Slant Correction in Script Classification. The IEEE International Conference on Signal and Image Processing Applications (ICSIPA’09), pp. 15-20.

Rehman, A. and Dzulkifli, M (2008). A Simple Segmentation Approach for Unconstrained Cursive Handwritten Words in Conjunction of Neural Network. International Journal of Image Processing, vol. 2 issue 3, pp. 29-35.

Rehman, A. and Saba, T. (2011a). Performance Analysis of Segmentation Approach for Cursive Handwritten Word Classification on Benchmark Database, Digital Signal Processing **, v**ol. 21, pp. 486-490

Rehman, A. and Saba, T. (2011b). Document Skew Estimation and Correction: Analysis of Techniques, Common problems and Possible Solutions. Applied Artificial Intelligence, vol. 25, No.9, pp. 769-787.

Rehman, A. Dzulkifli, M. and Kurniawan,F. (2008a). Line and Skew Removal from Off-line Cursive Handwritten Words. International Journal of Research (Science) Gomal University Pakistan, vol. 24, No.2, pp. 28-33.

Rehman, A. Kurniawan, F. and Dzulkifli M. (2008b). Off-line Cursive Handwriting Segmentation, A Heuristic Rule-based Approach. Journal of Institute of Mathematics and Computer Science (Computer Science Series) Kolkata, India, vol. 19(2), pp. 135-139.

Rehman, A. and Dzulkifli, M. (2009). Neuro-Heuristic Approach for Segmenting Cursive Handwritten Words. International Journal of Information Processing (IJIP), vol. 3(2), pp. 37-46.

Rehman,A. Saba, T. and Sulong, G. (2010). An Intelligent Approach to Image Denoising, Journal of Theoretical and Applied Information Technology, vol. 17(1), pp. 32-36

Romeo-Pakker, K., Miled, H., and Lecourtier, Y. (1995). A new approach for Latin/Arabic character segmentation. Proceedings of the Third International Conference on Document Analysis and Classification. 14-16 Aug 874-877.

Saba, T. Rehman, A. Elarbi-Boudihir, M. (2014). Methods And Strategies On Off-Line Cursive Touched Characters Segmentation: A Directional Review, Artificial Intelligence Review vol. 42 (4), pp. 1047-1066. doi 10.1007/s10462-011-9271-5.

Saba, T. Rehman, A. Altameem, A. Uddin, M. (2014) Annotated comparisons of proposed preprocessing techniques for script recognition, Neural Computing and Applications, vol. 25(6), pp. 1337-1347 , doi. 10.1007/s00521-014-1618-9.

Saba, T., Almazyad, A.S. Rehman, A. (2016) Online versus offline Arabic script classification, Neural Computing and Applications,vol.27(7), pp 1797–1804, doi. 10.1007/s00521-015-2001-1.

Saba,T., Rehman, A., Al-Dhelaan, A., Al-Rodhaan, M. (2014) Evaluation of current documents image denoising techniques: a comparative study , Applied Artificial Intelligence, vol.28 (9), pp. 879-887, doi. 10.1080/08839514.2014.954344.

Saba,T. Almazyad, A.S., Rehman, A. (2015) Language independent rule based classification of printed and handwritten text, IEEE International Conference on Evolving and Adaptive Intelligent Systems (EAIS), pp. 1-4, doi. 10.1109/EAIS.2015.7368806.

Saba T. Rehman, A. and Sulong, G. (2010b) .Non-Linear Segmentation of Touched Roman Characters Based on Genetic Algorithm, International Journal on Computer Science and Engineering, vol. 2(6), pp. 2167-2172.

Rehman, A. and Saba, T. (2012). Off-line cursive script recognition: current advances, comparisons and remaining problems, Artificial Intelligence Review, vol. 37(4), pp.261-268. doi. 10.1007/s10462-011-9229-
Saba, T. Rehman, A. and Elarbi-Boudihir, M. (2011b). Methods and Strategies on off-line Cursive Touched Characters Segmentation: A Directional Review. Artificial Intelligence Review, doi 10.1007/s10462-011-9271-5.

Saba, T. Rehman, A. and Sulong, G. (2010a). Cursive script segmentation with neural confidence, International Journal of Innovative Computing, Information and Control (IJICIC), vol.7(8), pp. 4955-4964.

Saba, T. Rehman, A. and Sulong, G. (2010c). Improved Statistical Features for Cursive Character Classification. International Journal of Innovative Computing, Information and Control (IJICIC), vol. 7(9), pp. 5211-5224.

Saba, T., Sulong, G. and Rehman, A. (2011a). Document Image Analysis: Issues, Comparison of Methods and Remaining Problems, Artificial Intelligence Review Springer, vol. 35(2), pp. 101-118, DOI: 10.1007/s10462-010-9186-6.

Journal of Engineering Technology
(ISSN: 0747-9964)
Volume 5, Issue 2, July, 2016, PP.156-168

Saba,T. Sulong, G. Rahim, S. and Rehman, A. (2010d). On the Segmentation of Multiple Touched Characters: A Heuristics Approach. Lecturer Notes in Computer Science (LNCS) Springer Verlag, pp. 540-544.

Saba, T. and Rehman, A. (2013) Effects of artificially intelligent tools on pattern recognition, International Journal of Machine Learning and Cybernetics, vol.4(2), pp 155–162.

Soleimanizadeh, S. Mohamad, D. Saba, T. Rehman, A. (2015) Recognition of partially occluded objects based on the three different color spaces (RGB, YCbCr, HSV) 3D Research, vol. 6 (3), 1-10., doi. 10.1007/s13319-015-0052-9

Sas, J., and Markowska-Kaczmar, U. (2007). Semi-Supervised Handwritten Word Segmentation Using Character Samples Similarity Maximization and Evolutionary Algorithm. 6th International Conference on Computer Information Systems and Industrial Management Applications, pp. 316-321.

Veloso, L.R., Sousa, R.P., De Carvalho, J.M. (2000). Morphological cursive word segmentation. Symposium on Computer Graphics and Image Processing, vol. 3(2), pp.337-343.

Verma, B. (2002). A contour character extraction approach in conjunction with a neural confidence fusion technique for the segmentation of handwriting classification. Proceedings of the 9th International Conference on Neural Information Processing, vol.5, pp. 2459-2463.

Verma, B. (2003). A Contour Code Feature Based Segmentation for Handwriting Classification. In Proceedings of the Seventh international Conference on Document Analysis and Classification(ICDAR), vol. 2, IEEE Computer Society, pp.1203-1221.

Verma, B., Gader, P. (2000). Fusion of multiple handwritten word classification techniques. Neural Networks for Signal Processing. Proceedings of IEEE Signal Processing Society Workshop, vol.2, pp.926-934.

Wiley and Sons, Inc.

Yanikoglu, B. and Sandon, P.A. (1998). Segmentation of offline cursive handwriting using linear programming, Pattern Classification, vol.31, pp.1825-1833.