International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2018 IJSRCSEIT | Volume 3 | Issue 3 | ISSN : 2456-3307

A Survey on Text Independent Writer Identification Methods

Anjumol Zachariah*[1], Anitha Abraham[2]

*[1]PG Scholar, Department of Computer science & Engineering, College Of Engineering, Kidangoor, Kerala, India

[2]Assistant Professor, Department of Computer science & Engineering, College Of Engineering, Kidangoor, Kerala, India

ABSTRACT

Writer identification based on handwriting, the behavioral characteristic of an individual, is an efficient method. Nowadays the most popular writer identification strategy is the text independent writer identification. This paper analyses the various text independent writer identification methods. There are some methods that combines both identification and verification systems. The most recent work in this area is implemented using recurrent neural network (RNN).

Keywords: Identification, RNN, Strokes, PDF

I. INTRODUCTION

Nowadays most electronic devices consist of digital surfaces that accept pen based inputs. This technological progress pushes forward the front tier for human-computer interaction. It is very important to label each individual uniquely. There are a lot of techniques that uses physiological characteristics such as fingerprint, face, iris, retina etc to identify an individual. Apart from these characteristics, handwriting is a behavioural characteristic to label an individual. Writer identification based on handwriting is an efficient strategy.

Writer identification is the process of determining the genuine writer from a list of candidates. According to the type of input data writer identification can be classified as either online or offline. In case of offline writer identification the image of the handwriting, samples are considered and it is processed to find out the author of the sample. The online writer identification considers the handwriting samples written on a digital surface using some input devices. It includes much richer information such as speed, angle, pressure etc. Writer identification can be implemented in the following two ways. Either in text-independent or text dependent manner. In text dependent writer identification the content of the text written by the writer is considered for the identification process. Higher accuracies can be achieved through this method. But, practically it is very difficult because it needs writing with fixed text content. The more general method is the text-independent writer identification in which the content of the text is not considered for the identification process. It is much suited for real application. However, it needs to capture the much richer information such as x, y co-ordinates, pressure, azimuth, altitude etc.

There are a lot of works based on text independent writer identification. It may be either online or offline. This paper analyses the various text independent writer identification methods. It ranges from a repetitive primitive's analysis system to the most recent system based on recurrent neural network (RNN). The difference between writer identification and verification is illustrated in figure 1.

The rest of this paper is organised as follows. Section II explains five text independent writer identification methods. Each of them gives a different approach to the identification process. Section III is the conclusion about all the works done so far. And finally references are given in Section IV.

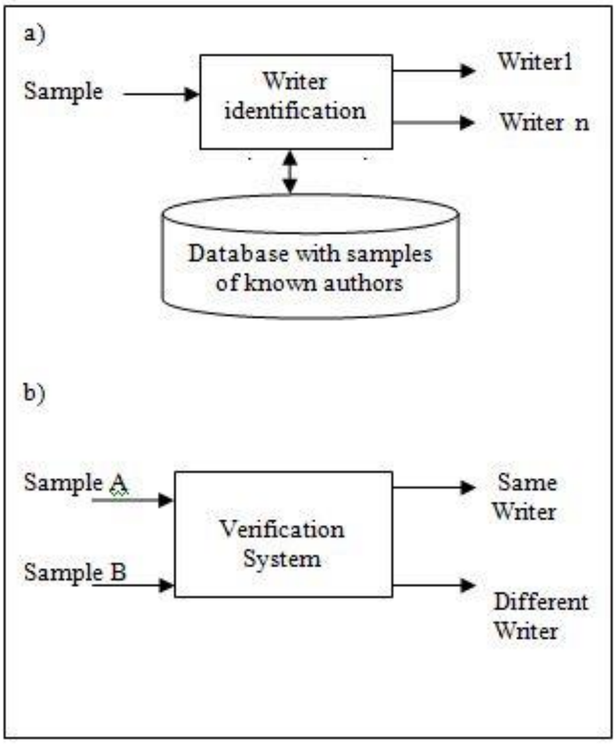

Figure 1. Writer Identification & Verification System

II. LITERATURE SURVEY

Here we analyses the various text independent writer identification methods.

A. Repetitive Primitive's Analysis

Anoop Namboodiri and Sachin Gupta in [1] proposes a text independent writer identification framework that uses a specified set of primitives of online handwritten data to authenticate the identity of the writer. It mainly concentrates on developing a secure and automatic person identification system. This work assumes that, in most of the cases people write uniquely and can be characterized based on the information present in their handwriting. They consider the variability in writing style as a major challenge. i.e., variability in style, shape, size etc. Variability in samples of a particular writer can increase further as the writing surfaces and conditions change.

The system mainly consists of two steps. First, it automatically identifies repetitive primitives in handwritten data of a particular script, and then uses the variations in those primitives to identify the writer. The identification primitives consist of the following steps.

a) Defining primitives
b) Extraction of primitives & representation
c) Calculation of similarity measures
d) Identify Consistent primitives
e) Writer identification

This framework allows us to learn the properties of the script and the writer simultaneously and hence can be used with multiple languages or scripts. This framework is applicable to any script where the specified primitives are present. The applicability of this technique include automated identity verification systems, such as ATMs and secure data access devices, where users can be authenticated based on a signature, name or password written using a stylus.

B. Textural and Allographic features identification

Marius Bulacu [2] proposes an effective technique for automatic writer identification and verification using probability distribution function (PDF) extracted from the handwriting images to identify the genuine writer. The technique work independent of the content of the text and it uses scanned images of handwriting samples for the processing. It operates at two levels textual level and character-shape level (allograph) level. The texture-level features are information regarding the pen-grip and preferred writing slant. While the allograph-level features indicate, the character shapes. It includes three main processing steps a) feature extraction b) feature matching/feature combination c) writer identification and verification.

This paper distinguishes between the writer identification and writer verification operations. The writer identification system performs a 1-to-many search in a large database with handwriting samples of known authorship and returns a likely list of candidates. The output of writer identification system will be analysed in detail by a human expert. Writer verification system involves a 1-to-1 comparison with a decision as to whether or not the two samples are written by the same person as shown in fig. The system involves minimal training. This study assumes that handwriting samples are produced using a natural writing attitude. Forged or disguised handwriting is not addressed here. The two processes are shown schematically in Figure 1.

C. Stroke's PDF

Bangyu Li et.al, [3] proposes a novel approach for online writer identification based on the stroke's probability distribution function (SPDF). The SPDF is used as the writer feature. Then it extracts four dynamic features to characterize writer individuality. It develops new distance measurement and combines dynamic features in reducing the number of characters required for online text-independent writer identification. In this paper with the information extracted from the strokes, an analysis of writing style is obtained. It select distribution of pressure, velocity, altitude and azimuth of i=12 primary stroke types for the representations of the writing, gestures and movements.

D. White Board Data Analysis

Andreas Schlapbach et. al, [4] proposes a writer identification using online handwriting captured from a white board. It is a text and language independent online writer identification system. The system is based on Gaussian Mixture Model (GMM) for the distribution on features extracted from the handwritten text. For n different writers it creates n different GMMs. It considers the context of a smart meeting room. The aim is to automate standard tasks usually performed by humans in a meeting. An important task in a smart meeting room is to capture the handwriting appear on a whiteboard during a meeting. This paper addresses the problem of identifying the author of a text written on a white board. i.e., it labels the handwriting with the writer's identity.

It includes both training and testing phases. The schematic overview of the two phases is shown in Fig 2. The training phase is used for the creation of one GMM for each writer. It includes the following two-step procedure. In the first step, all training data from all writers are used to train a single, writer independent universal background model (UBM). In the second step, for each writer a writer specific model is obtained by adaptation using the UBM and training data from that writer. As a result of the training procedure, we get a model for each writer. In the testing phase, a text of unknown identity is presented to each model. Each model returns a log-likelihood score, and these scores are sorted in descending order. Based on the resulting ranking, the text is assigned to the person whose model produces the highest log-likelihood score.

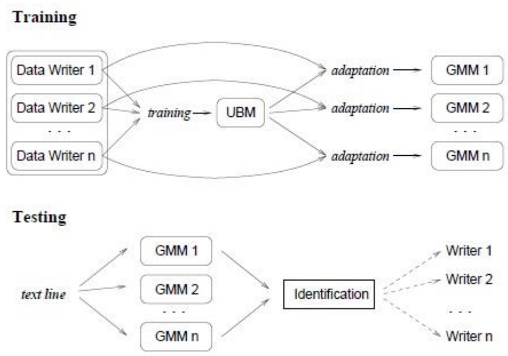

Figure 2. White Board Data Analysis Using GMM

E. Writer Identification Using RNN

Xu-Yao Zhang et. al, [5] proposes a writer identification using online handwriting captured using pen enabled input device. The purpose is to determine the genuine writer from a list of registered candidates according to the similarity between their handwriting, without considering the content of the text written by them. From the online handwriting information, a vector named random hybrid stroke (RHS) is generated. These RHSs are used to train the RNN and generate a particular score for each writer. During testing, the same process is repeated for another sample of handwriting and an ensemble-based decision is taken to find out the writer. The system is implemented using RNN with bidirectional long-short term memory (LSTM). The training model using bidirectional LSTM is shown in Figure 3.

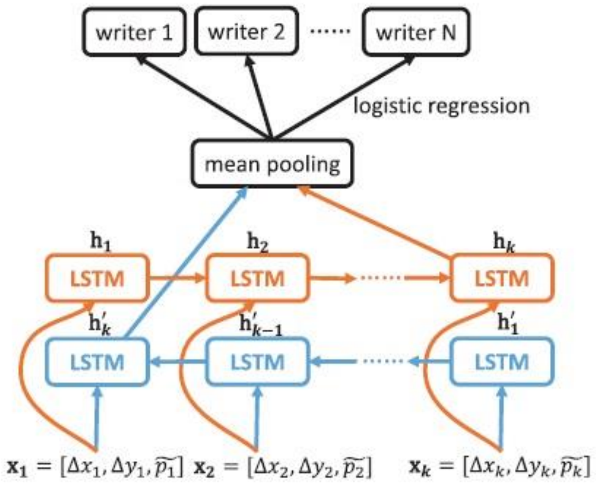

Figure 3. Bidirectional RNN Model for Writer Identification

It uses RNN for both training and testing. Initially the raw data collected is converted into a point-level representation and then a difference vector or stroke level vector is formed by taking the difference between adjacent vectors. One stroke is the basis unit used for segmentation. Multiple strokes are combined to form a RHS vector. The RNN training is done using the RHS vectors. After training a particular score is generated for each writer. During testing also the same process is repeated for the test sample and an ensemble-based decision is taken to find out the writer by comparing with the scores generated during training phase. Fig 4 gives an idea about the actual illustration of writer identification system using RNN. From the raw data , many RHSs are randomly sampled. After that, each RHS is fed into the RNN model individually to produce a probability histogram. At last all the histograms are averaged to make the ensemble-based decision.

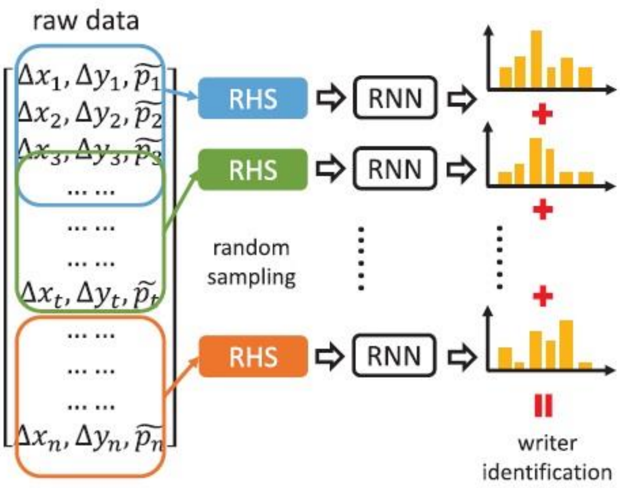

Figure 4. Illustration of writer identification using RNN

III. CONCLUSION

Writer identification is an important task and it finds its applications in the field of banking, digital forensic etc. As technology grows day by day the handwriting samples are more available online rather than offline. A lot of research work is done in the field of online handwriting identification. Most of the methods focus on identifying the author of a handwriting sample without considering the actual content which is written by him. This paper analyses five most recent works done in the area of text independent writer identification. Each method is briefly described and architectures are provided for some of the methods.

IV. REFERENCES

[1]. A.Namboodiri and S.Gupta."Text independent wrtier identification from online handwriting," Tenth International Workshop on Frontiers in Handwriting Recognition, pages 131-147, October 2006.

[2]. M.Bulacu and L.Schomaker, "Text-independent writer identification and verification using textural and allographic features," IEEE Trans.Pattern Anal.Mach.Intell., vol.29, no.4, pp.701-717, Apr.

[3]. B.Li, Z.Sun, and T.N.Tan."Online text-independent writer identification based on stroke's probability distribution function," Proc.of 2th ICB, pages 201-210, 2007.

[4]. A.Schlapbach, M.Liwicki, and H.Bunke, "A writer identification system for on-line whiteboard data," Pattern Recognit., vol.41, no.7, pp.2381-2397, 2008.

[5]. Xu-Yao Zhang, Guo-Sen Xie, Cheng-Lin Liu, and Yoshua Bengio, "End-to-End Online Writer Identification With Recurrent Neural Network", IEEE Trans.Human machine sym.2016.

[6]. Y.Yamazaki, T.Nagao, and N.Komatsu."Text-indicated writer verification using hidden markow model," Proc.of 7th ICDAR, pages 329-332, 2003.