International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169
Volume: 5 Issue: 6 232 – 235

Personnel Identification Using Handwriting, Tested On Indian Writers

Anish Raj[2], Apoorva Chaudhary[2]
Student[1] , Assistant professor[2]
Amity University, Haryana, India
rajanish4@gmail.com[1], apoorvadhaka07@gmail.com[2]

Abstract: Author identification is a method of distinguishing the author of a document using their handwriting. The expansion of machine engineering, computer science and pattern recognition fields owes greatly to one of the extremely challenged problem of handwriting identification. There are several ways of personnel identification like passwords, PIN. These give an extremely secure access to approved users, but credit cards are often purloined, whereas passwords and PIN are often forgotten or cracked. For this reason the biometric automatic identification of people supported by their physical or behavioural characteristics has gained widespread importance. Writing of an individual has some options that are distinctive to each person therefor are often used for identification. Scanned pictures of written documents are divided into words and these words are additionally divided into characters for word level and character level author identification. A collection of options are extracted from the metameric words and characters. The prominent feature that outperforms all others is that of the angle combination of 2 hinged edge fragments.[1]

The software test was conducted on English language handwriting of 30 different Indian people.

Keywords- *Feature extraction, Biometric, Statistical, Model based, Run length, Edge hinge, Graphemes.*

*****

1. Introduction

The significance and scope of writer identification is becoming more prominent in these days. Identification of a writer is highly essential in areas like forensic expert decision making systems, biometric authentication in information and network security, digital rights administration, document analysis systems and also as a strong tool for physiological identification purposes. Writer identification falls under the category of behavioral biometrics. Handwriting analysis is effectively implemented in the field of data retrieval either textually or graphically.

Writer identification mode can be generally classified into two types as online and offline. In online, the writing behavior is directly captured from the writer and converted to a sequence of signals using a transducer device but in offline the handwritten text is used for recognition by extracting information from scanned images. On-line writer identification is extensively considered as more challenging than off-line because it contains more information about the writing style of a person, such as pressure, speed, angle which is not available in the off-line mode.[5]

Writer identification approaches can be categorized into two types: text-dependent and text-independent methods. In text-dependent methods, a writer has to write the identical text to perform identification but in text independent methods any text may be used to establish the identity of writer.

There are basically two stages – feature extraction stage and classification stage. In the feature-extraction stage, features are extracted from handwriting and are stored in feature vectors. In the classification stage, the feature vectors are mapped onto classes representing the writers.

There are various approaches to build an algorithm for handwriting identification. Two such approaches are:

1. Statistical.
2. Model based.

These are described briefly in the following paragraphs:

2.1 Statistical approach

Research in automatic writer identification has mainly focused on the statistical approach. It involves extraction and specification of statistical features as in entropy,slant distribution,run length distributions and edge-hinge distribution.

2.1.1 Edge direction distribution:

In edge direction distribution, first the edge of the binary image is detected using Sobel detection method. Next, the first black pixel in an image is found and this pixel is considered as center pixel of the square neighborhood. Then the black edge is checked using logical AND operator in all direction starting from the center pixel and ending in any one of the edge in the square.[2] In order to avoid

redundancy the upper two quadrants in the neighborhood is checked because without on-line information, it is difficult to identify the way the writer travelled along the edge fragment. This will gives us "n" possible angles. Subsequently, the verified angles of each pixel are counted into n-bin histogram which is then normalized to a probability distribution which in turn gives the probability of an edge fragment oriented in the image at the angle measured from the horizontal.

2.1.2 Edge hinge distribution:

The edge-hinge distribution feature outperforms all other statistical features. Edge-hinge distribution is a feature that characterizes the changes in direction of a writing stroke in handwritten text. This distribution technique is extracted with the help of a window that is slid over an edge-detected binary handwritten picture. Whenever the central pixel of the window is on, the two edge fragments (i.e. connected sequences of pixels) emerging from this central pixel are considered. Their directions are measured and stored as pairs.[3] A joint probability distribution is obtained from a large sample of such pairs. The main limitation of the edge-hinge feature is that it only evaluates changes in direction on a single scale, rather than on multiple scales. For the edge-hinge feature, Bulacu et al. found an identification accuracy of about 63% on the Firemaker dataset using 250 distinctauthors.

2.1.3 Run length distribution:

Another feature is run length distribution, which are determined on the binarized image taking into consideration either the black pixels corresponding to the ink trace or, more beneficially, the white pixels corresponding to the background. [6] Whereas the statistical properties of the black runs mainly pertain to the ink width and some limited trace shape characteristics, the properties of the white runs are indicative of character placement statistics. There are two basic scanning methods: horizontal along the rows of the image and vertical along the columns of the image. As similar to the edge-based directional features described above, the run length histogram is normalized and interpreted as a probability distribution.[8]

2.2Model- based approach

This approach is based on a codebook made from models of graphemes. Graphemes are small strokes of handwriting, which are extracted by applying a robust segmentation algorithm on a handwriting image. Graphemes diff er from the edge fragments used for the construction of edge hinge distributions because of the used segmentation algorithm.

With a part of handwriting, grapheme features are extracted and matched to the grapheme models that are already present in the codebook. The matching is based on the Euclidean distance between the grapheme contours.[7] For each grapheme model, the number of successful matches is counted, yielding an approximation of a probability density function. In this approach, the writer is viewed as a grapheme generator that yields a characteristic probability density function. This function is taken as a feature for writer identification.

By combining graphemes and edge hinge distribution, Schomaker et al. foundauthor recognition accuracy of 97% on the Firemaker dataset with 150 distinct authors.[1]

2.3Classification stage:

Classification is performed with pdist function, which calculates the distance between vectors using various distance types (Manhattan distance, Euclidian distance or Chi-square distance).[4]

3. Methodology and software used:

Matlab was used to apply the algorithms and tests. The software implemented had the algorithm combination of edge hinge distribution and grapheme codebook (given by LVD Maaten), which was further improved to have a Graphic User Interface and the scalar value output to describe the variance in distance (Manhattan distance) between two most closely matched images. The images have to be converted into grayscale for extracting important features. The GUI had following options:

1. Selecting an image.
2. Adding the image to database.
3. Database information.
4. Identifying the two closest matching images.

The test was performed by the following validation method. First of all, a folder for sample handwriting images with 30 jpeg files was formed. Then a folder with the images to be tested was created (test folder), having similar 30 images of same writers but with different set of sentences and paragraphs. Then, one by one all the images from sample folder were fed into the software to train it and extract features, consequently the extracted information from each image was added into a database. Afterwards, the images from the test folder were tested for determining their closest match from the sample images in the database.

4. Results:

Whenever a handwriting image was tested, the software would respond with an output indicating the image name or

 ISSN: 2321-8169
232 – 235

number with the closest match and the variance in those two images, implying the images with least variance were most likely to be of the same person.

Having tested 30 different handwritings from Indian writers (in English), the variance between all the matched images varied from 0.085 to 0.100. And the success percentage of matched images was 100%.

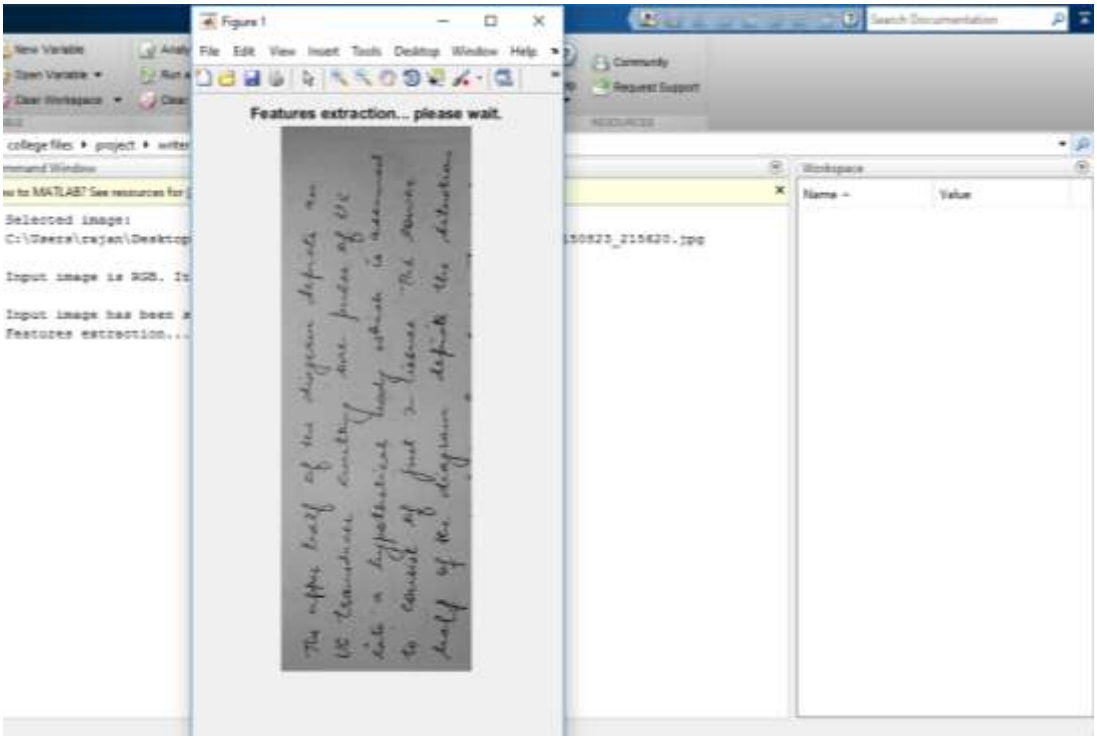

Fig. Feature extraction stage

Fig. Results

5. Conclusion:

The use of automatic and computation-intensive approaches in this application domain will allow for massive search in large databases, with less human intervention than is current practice. By reducing the size of a target set of writers, detailed manual and microscopic forensic analysis becomes feasible. It is important to note also the recent advances that have been made at the detailed allographic level, when character segmentation or retracing is performed by hand, followed by human classification. In the foreseeable future, the toolbox of the forensic expert will have been thoroughly modernized and extended. Besides their forensic applicability, the methods described here may have interesting potential applications in the field of historic document analysis. Examples are the identification of scribes on medieval manuscripts or identification of the printing house on historic prints. The only significant problem with handwriting based personal identification systems is that of aging, however it is expected to have a huge potential for research and development in the field of forensic handwriting analysis and writer identification. Particularly, the distributed client/server architecture allows for joining and sharing collected handwriting data, domain-specific knowledge, and implemented software routines. The current results obtained produced high accuracy levels, but more research is needed by including vast number of samples and by also including different handwritten languages in the database for checking the performance levels of the system.

References:

[1] M. Bulacu, L. Schomaker, and L. Vuurpijl.Writer identification using edge-based directional features. In Proceedings of ICDAR 2003, pages 937–941, Edinburgh, UK, 2003.

[2] L. Schomaker, M. Bulacu, and K. Franke. Automatic writer identification using fragmented connected-component contours. In Proceedings of the 9th IWFHR,pages185–190, Tokyo, Japan, 2004.

[3] Said, H., Tan, T., Baker, K. ―Personal Identification Based on Handwriting‖. Pattern Recognition Journal, vol. 33; pp 149-160, 2000.

[4] Marti, U., Messerli, R., Bunke, H. ―Writer Identification Using Text Line Based Features‖. 6th International Conference on Document Analysis and Recognition (ICDAR), Seattle (USA), pp 101-105, 2001.

[5] Srihari, S., Cha, S., Arora,H., Lee, S. ―Individuality of Handwriting : A Validity Study‖, 6th International Conference on Document Analysis and Recognition (ICDAR), Seattle (USA), pp 106-109, 2001.

[6] Al-Ma‘adeed S, Mohammed E, AlKassis D, Al-Muslih F, “Writer identification using edge-based directional probability distribution features for Arabic words,” IEEE/ACS International Conference on Computer Systems and Applications, pp.582-590, 2008.

[7] Bulacu M and Schomaker L, “Text-independent writer identification and verification using textural and allographic features,” IEEE Trans. on Pattern Analysis and Machine Intelligence, pp.701–717, April 2007

[8] B. Arazi. Handwriting identification by means of run-length measurements. IEEE Trans. Syst., Man and Cybernetics, SMC-7(12):878–881, 1977.