Applied GIS

ISSN: 1832-5505
Vol-5 Issue-01 Jan 2017

A Review of Gender prediction from handwriting using different algorithms

Payal patel[1], Abhishek Gupta[2]
[#]*School of Computer Science& Engineering, Shri Mata Vaishno Devi University, Kakryal, katra, Jammu & Kashmir, India*

***Abstract*—** The study of how to determine a person's gender based on their handwriting has recently gained interest. In forensic document examination, author identification, signature verification, etc., it has a number of intriguing uses. Predicting age, gender, country, etc., from a person's handwriting is a common use of handwriting recognition technology. More scrutiny is being given to gender prediction than any other application. One use of determining a person's gender based on their handwriting is creating psychological profiles of the authors. Numerous academics are exploring the possible gender differences in handwriting. The feature extraction stage is crucial to the whole predictive process. Various algorithms or techniques are used to extract information for use in a gender prediction system. This article provides an overview of the most common methods used to complete this difficult job. The first half of this article provides background information on the prediction issue at hand, while the second covers the methodology behind the gender prediction system's four stages, and the third offers a summary and last thoughts.

Keywords — **Keywords: Handwriting recognition (HWR), Gender prediction (GP), Image processing (IP), and Forensics (F).**

INTRODUCTION

Since the era before computers, identifying writers via their handwriting has been a common practice. Forensic scientists, document examiners, paleontologists, physiologists, etc. might all find this topic fascinating. Although handwriting analysis has been done manually for some time, recent technological and logarithmic advances in computer science have led to the rise in popularity of automated handwriting analysis utilizing computers. The most well-known use of handwriting analysis may be seen in the banking industry, where customers' signatures are checked against a database to ensure their identities. Although everyone learns how to write from books, each person eventually finds his or her unique voice when putting pen to paper. It's impossible for any two people to have identical fingerprints, and the same is true of their handwriting. While we all learned to write in a very uniform fashion at school, we each eventually developed a unique style that reflected who we were. In their study, Rizvi et al.[1] demonstrate how a person's personality may be deduced from their outward behavior and expression. Handwriting is a synthesis of the two. Research on handwriting uses a technique called graphology [2]. Among all the uses for handwriting analysis, handwriting recognition has the largest impact. Signature verification [3], writing style categorization [4], predicting other demographic factors (age, handedness, etc.), document verification/identification, etc. are all fascinating applications of handwriting analysis.

It was initially recognized by Binet in 1906, and has been reiterated many times since Hartley, 1991 [5] that a writer's gender may be inferred from his or her handwriting. Hayes [6] conducted five tests to test the hypothesis that male and female assessors reliably distinguish between the sex of authors. According to the results of his investigation, it is possible to infer a person's sex from their handwriting 75% of the time. According to Kozlowslu and Cutting [7], it is possible to distinguish between male and female based on their body motions with an accuracy of 60-70%. When their findings are combined, it's clear that people's handwriting and walking styles are highly correlated. Therefore, the act of writing may be seen as a sequence of human motions. They learn this procedure from their peers, and in the process, they reveal their gender. Many professionals in the fields of neuroscience and forensics have taken an interest in this notion. The science of graphology reveals a link between a person's health issues and the characteristics of their handwriting. Several physical disorders, as well as personality traits, are indicated by signs in a person's handwriting, as described in detail in articles [8]– [9], as illustrated in tables 1 and 2.

ISSN: 1832-5505
Vol-5 Issue-01 Jan 2017

TABLE I
HANDWRITING AND PERSONALITY CHARACTERISTICS

Handwriting characteristics	Personality characteristics
Spacing of writing	State of mind
Speed of writing	Amount of energy
Slope of writing	Degree of affection
Size of writing	Concentration powers
Pressure, etc	Emotional State

TABLE II
MEDICAL CONDITIONS AND HANDWRITING CHARACTERISTICS

Handwriting characteristics	Health conditions
Varying degree of pressure	Hypertension (High blood pressure)
So small and cramped	Parkinson's disease
Altered handwriting and irregular letters	Alzheimer's disease
Writing in a more angular fashion.	Arthritis

It is believed that handwriting indicates more than 5000 different aspects of your personality. Among various attributes, it has shown that gender is more significantly affecting the writing styles of a person. The difference in writing styles [10] of male and female is due to hormones in them [5] - [6] - [11]. Some of the reasons are:

A. Biological attributes

- Structure of hand
- Handedness (left or right)
- Strength of muscles

B. Memetic (cultural) attributes

- Styles of pen grip
- drawing character shape

ISSN: 1832-5505
Vol-5 Issue-01 Jan 2017

TABLE III COMPARISON TABLE

Author name and year	Preprocessing Technique Used	Features Extracted	Classification method Used	Dataset used	Experimental Results
Gattal et al.(2018) [12]	Global thresholding	Textural Features	SVM	QUWI dataset	79.50%
Ahmed et al.(2017) [13]	None	Geometric features	SVM ,DT, RF, ANN and ensemble classifiers	QUWI dataset	67%
Mirza et al.(2016) [14]	None	Gabor features	ANN	QUWI dataset	70%
Bouadjenel et al.(2015) [15]	Grid is applied to an input image	Gradients Features	SVM	IAM dataset. KHATT dataset	76%
Maadeed et al.(2014) [4]	Binarizing the data	Geometric features	Random forest (RF) and kernel discriminative analysis(KDA)	QUWI dataset	KDA-73.6% RF- 74.8%
Siddiqi et al.(2014) [16]	None	Global and textural features	ANN SVM	QUWI dataset. MSHD dataset	SVM- 73.02% ANN- 69.44%

With the fast adaptation of computers with increase in technologies, this research is becoming more automated than manual handwriting analysis. Various techniques were developed related to it, which extracts necessary (features) information from scanned images and is employed using various classification methods. In this section various research studies are compared and in the next section these different techniques are discussed in detail which can be used by author for their research work.

RELATED WORKThe basic steps of a gender recognition system are:

A. Image acquisition
B. Image preprocessing
C. Features Extraction
D. Classification

A. *Image acquisition*

To start handwriting recognition, the first and the most important thing needed is the images on which the techniques is applied for further processes. The images used can be online or offline images. Jaegar in [17] made a online handwriting recognition system using online images. The difference between online and offline images lie in the process of its acquisition. In online images, a recording system records the movement of pen, pen grip and direction. Whereas in offline images, the handwritten data on a paper is scanned using OCR (optical character reader) [18].

B. Image preprocessing

Image preprocessing is the technique that is applied on the intensity images to improve the quality of the images. It is for the improvement of the images and to suppress the unwanted distortions to boost the important features so that the quality of research is improved in efficiency. Various steps involved in image pre-processing are:

1) Converting RGB into gray image: RGB image is converted into gray scale image in which all three (red, green and blue) colors have equal intensities as a result less inormation needs to be provided for each pixel.

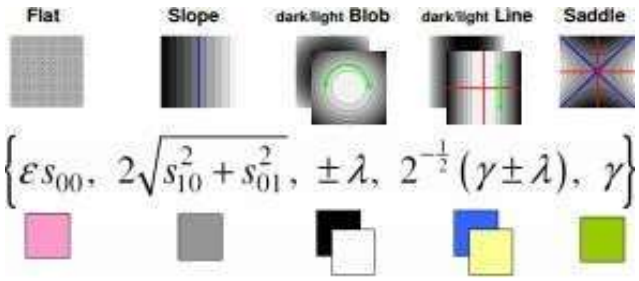
Converting grayscale into binary image: Images are converted from grayscale to binary images or are binarized using Otsu'sthresholding [19].

2) Filtering or smoothing: this is done to enhance the quality of image by removing the noise from background.

B. Features Extraction

It means to transform the input data into features set. The important properties of input data is known as features are extracted whichhelps in classifying the input patterns. There are various for features extraction. Some of them are:

Oriented Basic image features: OBIF is a textural descriptor and is the extension of the Basic Image Features [20] - [21]. BIF tag the image configuration as one of the seven local types of symmetry which are, flat, slope-like, maxima-like, minima-like, dark line-like, light line-like and saddle-like. Pink, gray, black, white, blue, yellow and green colors respectively are used to picture these symmetry types as shown in Fig. 1. Labels are allocated as the maximal of one of the seven symmetry expressions in the 2nd order derivatives. These expressions derivatives are calculated as the inner product of their 2nd order Gaussian derivative and Image [21]. The inner



product is,

ISSN: 1832-5505
Vol-5 Issue-01 Jan 2017

$\lambda = \sigma^2 (s_{20} + s_{02})$, $\gamma = \sigma^2 ((s_{20} - s_{02})^2 + 4s^2_{11})^{1/2}$

Fig.1 Basic Image Features [20]

In oBIF, the orientation is augmented with BIF in a controlled manner. Pink, black and white (flat, maxima and minima) does not have any orientation and so remain unoriented. Gray BIF is augmented with gradient orientation of 8-quantised and Blue, yellow, and green (dark/light lines and saddles) is augmented using principal orientations of 4-quantised. There are several encoding schemes suchas SIFT [22].

2) Orientation and curvature: The contours of the writing provide the information about the slant and curvature using the set of features computed from the writing.

- *Polygon-based feature:* This feature is computed using polygonization algorithm [23]. The contours of the words are estimated by a set of line segments (polygon). The slope of each line segment and for curvature of a writing and angle between each pair of neighboring line segment is calculated and there distributions are used as a feature. The slope of the line segment is quantized into 8-bins of the histogram from -90^0 to 90^0. The angle (curvature) is computed as,

$$\alpha_i = \pi - \arccos \frac{V_i . V_{i+1}}{|V_i| |V_{i+1}|}$$

where V_i and V_{i+1} being the vectors from (x_{i-1},y_{i-1}) to (x_i,y_i) and from (x_i,y_i) to (x_{i+1},y_{i+1}). The angle is also computed in 8-bins histogram from (0^0 to 180^0).

3) Texture-based features: Texture based features is associated with the overall representation of the writing. It represents the texture of the writing that is the feel of the writing. One method used for this is:

- *Local Binary Pattern:* This algorithm was firstly introduced by Ojala [24] and is used many times for writer identification [25][26]. Here, the set of neighboring pixels (V_0 to V_8) is compared with its central pixel and to generate a binary pattern as follows:

The adjacent pixels is compared with its central pixel (V_0 being the central pixel), if $V_i < V_0$ then assign 0 to the pixel else 1 to that pixel. This will generate the binary pattern. The resulting 0's and 1's number is consider as a binary number. as a set of training data, and as an input it takes the set of input data to predict them in one of the distinct classes (male and female). This performs classification and training using kernel function [33] - [34]. It outputs the maximum marginal hyperplane to differentiate between the two classes. As a result, the data item which is in the either class belongs to that class only.

(2).To generate the LBP code, the corresponding binary number is multiplied with its weight and all are added to produce a LBP code for a central pixel. The histogram of LBP code is used as a descriptor to characterize the texture.

A. CLASSIFICATION

In classification step, features extracted as mentioned above are used to decide which category the writer belongs to. The features are used as an input to the classifier and the output of the classifier is the desired result. There are numerous numbers of classifiers used but two mostly used are listed below:

I. Artificial Neural Network: The ANN is based on the biological neural network [27]. The ANN [28] is a 3 layer network, input layer, hidden layer and the output layer. The input layer has same number of neurons equal to that of the dimensionality of the feature vector. Output layer has 2 neurons (here, male and female) for binary classification [29] and the purpose of the hidden layer neurons is to determine the dimensionality of the input layer. A threshold activation function [30] is used to specify the output of a neuron in the output layer. The network is trained using dataset images, either using supervised learning or unsupervised learning [31] and then the rest of the images is tested on this trained classifier to produced the required result. The efficiency is clearly depends upon the training part of the method. (1).

II. Support Vector Machine: SVM is a supervised learning binary classifier. It is use for both classification and regression techniques [32]. This classifier is trained using features extracted such as direction, curvature or slope feature of the writing that is use

DISCUSSION

Lots of experiments has been take place in this search area by various scientists or scholars, and the experiments show that this problem of finding writer's sex from their handwriting is a complex problem. Hayes in 1996 [6] said that even small amount of material can produce sex judgments above chance level, even a single letter or a single pattern drawn by a writer is enough. Writing with the non preferred hand does not give any knowledge about the writer's gender. As it do not represent the true personality of a person. Also, other demographic traits such as age, is not fetched much easily by handwriting as gender prediction. Further manual experiments show that the women perform well than men in predicting the handwriting of the writer as male or female.

CONCLUSION

This research problem is one of the most researched one, several methods or techniques have been surveyed for it

ISSN: 1832-5505
Vol-5 Issue-01 Jan 2017

and the one recently used in the past few years with their accuracy rates were listed in the table 3. This problem has four steps, but the most important is the feature extraction, as this step is the base of the result in the form of training of the classifier. If the training is not done in proper manner, the result produced is not accurate. Among all features extracted the oBIF features shows significantly higher accuracy rate than all others, because it make use of the gradient values that is the contours value which surely defines the shape of the image. But as the machine does not work even if small part of a machine is not working properly, similarly an effective gender prediction system can be developed only by combining all steps properly.

REFERENCES

[1] F. Rizvi, "Personality Prediction Through Offline Handwriting Analysis," Journal of Multidisciplinary Engineering Science Studies, volume 3, issue 1, pages 1282-1288, 2017.

[2] "(a) History of Graphology," by T. Second, W. Study, C. Script, and R. Difference, no. d.

[3] Reference: [3] "Offline signature verification using classifier combination of HOG and LBP features," 2011 International Joint Conference on Biometrics, IJCB 2011, no. October 2015, 2011.

[4] For example, see [4]S. Al Maadeed and A. Hassaine, "Automatic prediction of age, gender, and nationality in offline handwriting," Eurasip Journal on Image and Video Processing, pp. 1-10, 2014.

[5] [5] J. Hartley, "Sex Differences in Handwriting: a Comment on Spear," Journal of Handwriting Research, vol. 17, no. 2, pages 141-145, 1991.

[6] As cited in [6] W. N. Hayes, "IDENTIFYING SEX FROM HANDWRITING," Perceptual and Motor Skills, vol. 83, pp. 791-800, 1996.

[7] &CUTTING, [7]. In 1977, E. Kozowsw, L. T., "Recognizing_Sex," appeared in Percepttori ar.d P.rychophysics, volume 21, issue 6, pages 575-580.

[8] [8] "https://dopasolution.com/what-your-handwriting-says-about-your-physical-health/."

[9] (9) "https://blog.whsmith.co.uk/what-does-your-handwriting-say-about-you/."

[10] 2007; pp. 207-208 in K. M. Koppenhaver's "Forensic document examination:principles and practice."

[11] Do variations in sex hormones influence handwriting style?," (J. R. Beech and I. C. Mackintosh, 2011). Evidence from digit ratio and sex role identification as factors of the sex of handwriting," Journal of Research in Gender and Development, 39, 459-468.

[12] According to [12] "Gender classification from offline multi-script handwriting images using oriented Basic Image Features (oBIFs)," published in Expert Systems with Applications, volume 99, issue February 2018, pages 155-167, 2018.

[13] According to [13] "Improving handwriting based gender classification using ensemble classifiers," M. Ahmed, A. Ghulam, H. Afzal, and I. Siddiqi, Expert Systems With Applications, volume 85, pages 158-168, 2017.

[14] "Gender classification from offline handwriting images using textural features," by A. Mirza, M. Moetesum, I. Siddiqi, and C. Djeddi; published in Proceedings of the 2017 International Conference on the Frontiers of Handwriting Recognition; pages 395-398.

[15] "Age, Gender, and Handedness Prediction from Handwriting using Gradient Features," by H. N. and Y. C. Nesrine Bouadjenek, pp. 1116-1120, 2015.

[16] Reference: [16] "Automatic analysis of handwriting for gender classification," by I. Siddiqi, C. Djeddi, and A. Raza, 2014.

[17] [17] S. Jaeger, S. Jaeger, S. Manke, J. Reichert, and A. Waibel, "Online handwriting recognition : the NPen ++ recognizer. Int J Doc Anal Recognit 3 : 169-180," no. May 2017, pp. 169-180, 2001.

[18] "Optical Character Recognition Using Image Processing," by S. G. Dafe and S. S. Chavhan, 2018, pp. 962-964.

[19] "Reviewing Otsu's Method For Image Thresholding," International Journal of Applied Engineering Research, vol. 10, no. August 2016, pp. 21777-21783, 2015. [19] D. S. T. P. Sunil L. Bangare, Amruta Dubal, Pallavi S. Bangare.

[20] Indian Council of Philosophical Research (2008), "Novel Image Feature Alphabets for Object Recognition," M. Lillholm and L. Griffin.

[21] [21] Griffin, L. D., and Lillholm, M., "Symmetry-sensitivities of derivative-of-Gaussian Filters," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 32, no. 6, pages 1072-1083, 2010.

[22] In 2004, D. G. Lowe published "Distinctive Image Features from Scale-Invariant Keypoints," which covers the topic in detail.

[23] Reference: [23] K. Wall and P. E. Danielsson, "A fast sequential method for polygonal approximation of digitized curves.," Computer Vision, Graphics, & Image Processing, vol. 28, no. 2, pages 220-227, 1984.

[24] "Performance evaluation of texture measures with classification based on Kullback discrimination of distributions," [24] T. Ojala, M. Pietikäinen, and D. Harwood.

[25] Pattern Recognition: Third International Conference Proceedings, pages 582-585, 1994.

[26] Reference: [25] A. Hassane, S. Al-Maadeed, and A. Bouridane, "A set of geometrical features for writer

ISSN: 1832-5505
Vol-5 Issue-01 Jan 2017

identification," Lecture Notes in Computer Science (containing subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 7667 LNCS, no. PART 5, pp. 584-591, 2012.

[27] According to [26] "Novel geometric features for off-line writer identification," by S. Al-maadeed, A. Hassaine, A. Bouridane, and M. A. Tahir, published in Pattern Analysis and Applications, volume 19, issue 3, pages 699-708, 2016.

[28] "Functional model of biological neural networks," by J. T. H. Lo, "Cognitive Neurodynamics," volume 4, issue 4, pages 295-313, 2010.

[29] "Research Paper on Basic of Artificial Neural Network," by M. S. B. Maind and M. P. Wankar, was published in 2014 in International Journal on Recent and Innovation Trends in Computing and Communication, volume 2, issue 1.

[30] "Machine Learning: A Review on Binary Classification," by R. Kumar and S. K. Srivastava, was published in 2017 in the International Journal of Computer Applications, volume 160, issue 7.

[31] For example, see [30] B. Karlik, "Performance analysis of various activation functions in generalized MLP architectures of neural networks," International Journal of Artificial Intelligence and Expert Systems, volume 1, issue 4, pages 111-122, 2015.

[32] The following is an excerpt from "Comparison of Supervised and Unsupervised Learning Algorithms for Pattern Classification," written by R. Sathya and A. Abraham and published in the International Journal of Advanced Research in Artificial Intelligence, volume 2, issue 2, pages 34-38, 2013.

[33] "PER-08: A Comparative Study of Classification and Regression Algorithms for Modelling Students' Academic Performance," P. Strecht, L. Cruz, C. Soares, J. Mendes-Moreira, and R. Abreu, Portugal, 2015, pp. 392-395.

[34] "Research on Kernel Function of Support Vector Machine # (1.," vol. 0, no. 4112045, pp. 1-6, 1986, L. Lijuan, S. Bo, and W. Xing.

[35] For example, see [34] H. Huanrui, "New Mixed Kernel Functions of SVM Used in Pattern Recognition," Cybernetics and Information Technologies, volume 16, issue 5, pages 5-14, 2016.