

# Identifying Decisive Features for Distinctive Analysis of Writings in Malayalam

Sreeraj.M, Sumam Mary Idicula

**Abstract**— This paper presents a writer identification scheme for Malayalam documents. As the accomplishment rate of a scheme is highly dependent on the features extracted from the documents, the process of feature selection and extraction is highly relevant. The paper describes a set of novel features exclusively for Malayalam language. The features were studied in detail which resulted in a comparative study of all the features. The features are fused to form the feature vector or knowledge vector. This knowledge vector is then used in all the phases of the writer identification scheme. The scheme has been tested on a test bed of 280 writers of which 50 writers having only one page, 215 writers with at least 2 pages and 15 writers with at least 4 pages. To perform a comparative evaluation of the scheme the test is conducted using WD-LBP method also. A recognition rate of around 95% was obtained for the proposed approach.

**Index Terms**— Feature extraction, Loop features, Distance features, Directional features, Writer identification, Malayalam.

## I. INTRODUCTION

THE significance and scope of writer identification is becoming more prominent these days. The writer identification techniques is currently used in wide arenas like the digital rights administration, forensic expert decision-making systems, document analysis systems and also as a strong tool for physiological identification purposes. The writer identification is also utilized in authentication system by combining with the writer verification for many fields of confidential data handling purposes. As a consequence, the number of researchers involved in this challenging problem is increasing immensely. A distinguished work in this course is [1] by Srihari et al which proposes the features necessary for the writer identification scheme at different levels. However there have been immense writer identification schemes available in languages like Chinese, Arabic, and Persian. There exists no writer identification scheme in Malayalam language up till date to the best of our knowledge.

Manuscript received February 28, 2011.

Sreeraj. M is with the Department of Computer Science, Cochin University of Science and Technology, Cochin, India (phone: +91-0484-2576368; e-mail: msreeraj@cusat.ac.in).

Sumam Mary Idicula is with the Department of Computer Science, Cochin University of Science and Technology, Cochin, India (e-mail: sumam@cusat.ac.in).

The motivation for Malayalam writer identification scheme stems out from the following challenges posed by the language as well as some other factors.

(i) *Meager Allographic Variation of writers in Malayalam*

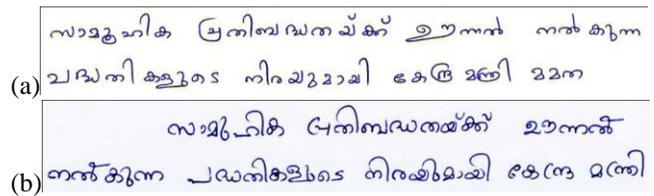


Fig. 1. Allographic variation of two writers in Malayalam: (a) extracted from writer1; (b) extracted from writer 2.

*language* – A major factor causing variation in handwriting is Allographic variation. Writer specific character shapes are derived from this variation. They are a threat to the automatic script recognition. In spite of this, it substitutes vital information for writer identification. This variation is very low in Malayalam handwriting when compared with other languages as shown in the two samples given in Fig 1.

(ii) *Insufficient discriminating capacity of a single character in Malayalam language* - A single character does not provide sufficient discriminating values and hence a combination of characters may be necessary to give out the prominent feature vector of the handwriting. This again adds the identification complexity.

(iii) *Non-existence of uppercase and lowercase in Malayalam language* – Writers’ adopts certain prominent styles related to upper case and lower case characters. Since Malayalam scripts do not have upper and lower cases this prominent discrimination cannot be applied. Same is the case with cursive style and Malayalam has no cursive form of writing.

(iv) *Absence of dataset*- Absence of the dataset of handwritten pages of different users in Malayalam pose a great challenge. Hence, a collection of handwritings of different users of similar as well as different data had to be collected for the purpose of implementation.

(v) *Writing impression* - Pen grip, the orientation of the wrist and the fingers together constitute a habitual parameter slant (shear) in the writing style of each user [2]. Malayalam scripts mainly contain loops and curves wherein every variation has to be considered. When different writings are compared this parameter is low. So there is a need of observing the minute changes in affine transform in the loops as well as the curves of each character in Malayalam script.

In this paper, Section II presents related work and taxonomy of the writer identification at different feature level. Section III portrays a design of the writer identifications scheme for Malayalam language. Section IV describes the features used for the purpose of writer identification. Section V explains the implementation details. Section VI outlines the results obtained and provides valid conclusions regarding the features selected for Malayalam documents. The paper is concluded in Section VII.

## II. RELATED WORK

A comprehensive review covering the research work in automatic writer identification until 1989 is given in [3]. An extension including work until 1993 has been published in [4]. The on-line handwritten data contains more information about the writing style of a person such as speed, angle, or pressure that is not available in the offline data. Thus, the online classification task is considered to be less difficult than the offline ones [5] [6]. Further, writer identification can also be divided into two parts: text-dependent and text-independent writer identification. This classification is mainly on the different feature level such as character, word, line, paragraph and the document level. Srihari et al. [1] [7] [8] propose a large number of features based on two categories such as Macro level and micro level operate at document/paragraph/word level of CEDAR database using multi-layer perceptron and achieved 98% accuracy. Zois and Anastassopoulos [9] perform writer identification and verification using single words based on the morphological operators. Zhang et al. [10] used the Gradient (192bits), structural (192 bits), and concavity (128 bits) features for the writer identification based on the k-nearestneighbor classification and yield 97.71%. Al-Ma'adeed et al. [11] employed edge-based directional probability distributions, combined with moment invariants and structural word features for the writer identification of Arabic word.

Said et al. [12], [13] propose a text-independent approach and derive writer-specific texture features using multichannel Gabor filtering and gray-scale co-occurrence matrices. A similar approach has also been used on machine print documents for script [14] and font [15] identification. But Bensefia et al. [16] [17] [18] [19] use graphemes generated by a handwriting segmentation method to encode the individual characteristics of handwriting independent of the text content. Bulacu et al. presented text-independent Arabic writer identification by combining some textural and allographic features [20] [21].

A multi-character level decision combination module for the off-line Chinese writer identification scheme is given by [50] which showed an accuracy of 88.41% for Top 10. A recent trend in writer identification scheme through subspace is proposed in [51]. This method reduces the elapsed time on computation of the identification process.

Fig.2 depicts the taxonomy of the writer identification at different feature level.

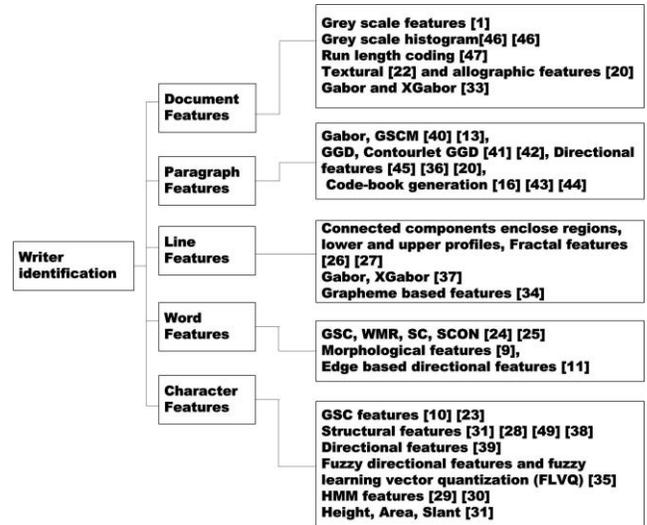


Fig. 2. Taxonomy of writer identification at various feature level.

## III. WRITER IDENTIFICATION SCHEME

Before discussing the scheme of writer identification for Malayalam language, there are some explicit features of the Malayalam script which require mention. They are outlined as below.

- Malayalam scripts are curvaceous in nature. It contains loops and curves. Normally they have low aspect ratio (ie. Height / Width ratio of consonant characters) which includes vowels, consonants, chillu , anuswaram, visargam, chandrakkala, consonant signs, left vowel signs , right vowel signs, conjunct consonants. This requires special mention here because the feature identification method adopted here separates each character on the basis of stroke made by the user.

- Two prominent ways of writing Malayalam scripts exists today. One followed by older generation and the other followed by younger generation. This will be helpful for first level of clustering the writers which they belongs. Some people belonging to older generation do rarely exhibit the habit of writing two or three characters connected.

- Since Malayalam script is an alphasyllabary of the Brahmic family they are written from left to right.

### A. Architecture

Any writer identification scheme generally consists of training phase, labeling phase and identification phase. The system architecture given in Fig.3 depicts in detail the feature extraction module of the training phase. The success rate of a writer identification scheme is highly dependent on the selection of an appropriate feature vector. When a feature vector is being selected, all the above mentioned characteristics of Malayalam script also has to be taken into account for. So we gave prime importance on the feature vector formed from a Malayalam script. The features which we are making use of in the scheme are described in Section IV in detail. The resultant feature vectors are fused together to

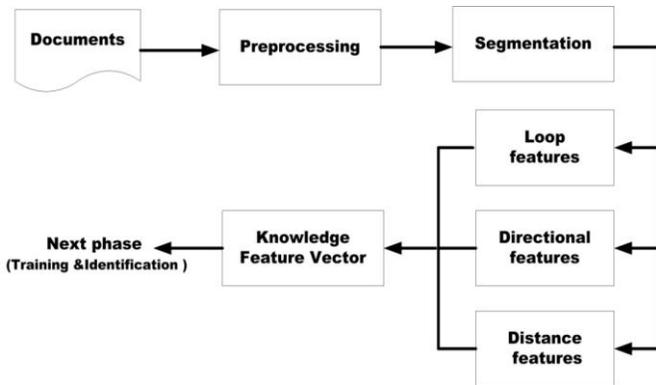


Fig. 3. System Architecture.

obtain the knowledge vector (which account for individuality) of each writer. Also this vector is used for both training and identification phases.

IV. FEATURES FOR WRITER IDENTIFICATION

Due to the low allographic variation in Malayalam handwritings, the writer identification scheme needs a prominent feature vector for automatic writer identification. It has also to be noted that Malayalam is a language that contains loops in almost eighty percentages of the characters. As of that, characters which don't form a loop cannot be so decisive in the feature vector. Also certain characters of a writer don't give an individuality to help for the identification. Prominent features can be observed in the hooks made by the user so dehooking is avoided in the automatic writer identification. Also breakage in loops is taken care by means of dilation. The features to be extracted for writer identification can be generally classified as: Character & Grapheme Level. Here we are making use of only the character level features.

**Character level:** In this level the loop features, directional features and distance features are considered for the purpose of forming feature vector.

A. Loop features

Loop features are the features that pertain to the characteristics of a loop including loop area, loop radius and loop roundness [48]. Since most of the Malayalam characters are circular in shape the loops and curves had to be dealt with minutely which provides essential information for distinguishing the different writing styles. After the study it has been concluded that a writer maintains his own style of loops and curves throughout his writings. The characteristics of loop features are as follows.

**Loop roundness (f1):**

It is observed that every writer's character that consists a loop maintains his own shape of roundness throughout his writing. Loop roundness measures its similarity to a perfect circle. The index of dissimilarity,  $d$  can be mathematically computed as [49].

$$d = \frac{d'}{C.radius} \tag{1}$$

Where  $d'$  is computed as:

$$d' = \frac{\sum_{p \in LC} \left( \sqrt{(p.row - C.row)^2 + (p.col - C.col)^2} - C.radius \right)^2}{n(LC)} \tag{2}$$

Where  $n(LC)$  is the count of the pixels at the edges of loop.  $C.row$  and  $C.col$  are the centre of the ideal circle and its radius is  $C.radius$ . The circle centre is estimated at the centre of gravity of the loop. It is calculated as

$$C.row = \frac{\sum_{p \in LOOP} p.row}{n(LOOP)} \tag{3}$$

$$C.col = \frac{\sum_{p \in LOOP} p.col}{n(LOOP)} \tag{4}$$

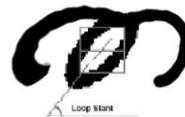


Fig. 4. Loop slant in the Malayalam letter 'tha'.

Where  $n(LOOP)$  is the count of pixels that belong to the loop.

**Loop Slant (f2):**

Each user has a habitual parameter of applying his own style of slant which would repeatedly see among his characters. The upper and the lower part of equal height of the loop have to be considered and the centre of gravity of both parts is calculated. Loop slant is measured as the angle of line connecting the centre of gravity of each part as shown in Fig.4.

**Relative Height and Width/Height ratio of loops (f3):**

The ratio between the height of the loop and the total height of the letter is the **Relative height**. It is mostly seen that this ratio remains consistent throughout one's writing. It is calculated using (5) and is shown in Fig.5.

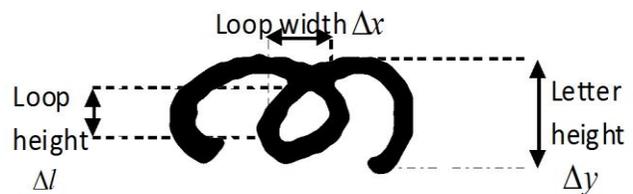


Fig.5. Relative height ratio in the Malayalam letter 'tha'.

$$Relative\ height\ ratio = \frac{\Delta y\ of\ the\ loop}{\Delta y\ of\ the\ letter} \tag{5}$$

The maximum and minimum 'x' values of loop and maximum and minimum 'y' values are calculated and then the difference between maximum and minimum values gives  $\Delta x$  and  $\Delta y$  respectively. The **Width/Height ratio** is calculated as

$$\text{Ratio} = \frac{\Delta x \text{ of the loop}}{\Delta y \text{ of the loop}} \quad (6)$$

### B. Directional features

#### Direction angle of the loop (f4):

This is used for distinguishing the broad and narrow loops of the writers. It is calculated as follows:-

Step1. Check whether the loop is ascender or a descender.

Step2. If the loop is an ascender loop, then using (7) calculate the angle between the x-axis and the vector between the intersection point and the highest point in the loop. It has a direction value between 0 and 180 degrees.

Else, using (7) calculate the angle between the x-axis and the vector between the intersection point and the lowest point in the loop. It has a direction value between 180 and 360 degrees.

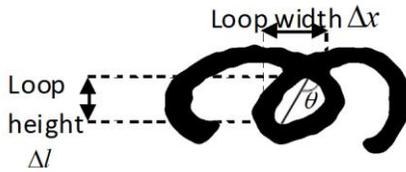


Fig. 6. Direction angle of the loop in the Malayalam letter 'tha'.

Step3. Calculate the average by adding up of these angles towards each pixel in the loop and by dividing the result by the number of angles.

Step4. Take the standard deviation of the average direction to quantify how much the loop direction differs from their mean. Following Fig.6. depicts the direction angle of the loop.

$$\text{Direction angle } (\theta) = \arctan \frac{\Delta y \text{ of the loop}}{\Delta x \text{ of the loop}} \quad (7)$$

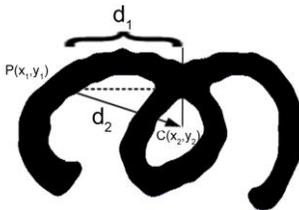


Fig. 7. Direction angle of the letter 'tha' in Malayalam.

#### Direction angle of the character (f5):

It is the angle subtended by the x-axis of the points in the stroke and the centroid of the character as shown in the Fig. 7. It is calculated as follows:

Step1. Calculate the distance  $d_1$  and  $d_2$  as shown in the

Fig.7.

Step2. Calculate the angle using (8)

$$\text{Angle } (\phi) = \cos^{-1} \frac{d_2}{d_1} \quad (8)$$

#### Slant of a character (f6):

It is the angle each character forms against the baseline. It is estimated on structural features by maxima and minima of the character are detected and targets uniform slant angle estimation as described in [52]. Another approach, is based on the projection histogram is applied as described in [53].

#### Curvature of the character (f7):

This feature is attained using the point based method and contour based method.

##### Point based method.

Step1. Take n equal points in the stroke of each character.

Step2. Calculate the curvature at each point Using the normalized first derivatives of each point using (9) and (10) and second derivatives are calculated by replacing

$x$  and  $y$  with  $\hat{x}$  and  $\hat{y}$  in and normalized using (10).

$$x'_i = \frac{\sum_{i=1}^2 i.(x_{i+1} - x_{i-1})}{2.\sum_{i=1}^2 i^2} \quad y'_i = \frac{\sum_{i=1}^2 i.(y_{i+1} - y_{i-1})}{2.\sum_{i=1}^2 i^2} \quad (9)$$

$$\hat{x}_i = \frac{x'_i}{\sqrt{{x'_i}^2 + {y'_i}^2}}; \quad \hat{y}_i = \frac{y'_i}{\sqrt{{x'_i}^2 + {y'_i}^2}} \quad (10)$$

Step3. Curvature at a point is the inverse of the radius of the osculating circle is calculated using (11).

$$K = \frac{\hat{x}' \hat{y}'' - \hat{x}'' \hat{y}'}{\left( \hat{x}'^2 + \hat{y}'^2 \right)^{3/2}} \quad (11)$$

##### Contour based method:

For a closed contour, circular convolution can be applied directly to smoothen the contour. For an open contour, however, a certain number of points should be symmetrically compensated at both ends of the contour when it is smoothed. The contour convolved with the Gaussian smoothing kernel. The curvature value of each pixel of the contour is computed using

$$K_i^j = \frac{\Delta x_i^j \Delta^2 y_i^j - \Delta^2 x_i^j \Delta y_i^j}{\left[ (\Delta x_i^j)^2 + (\Delta y_i^j)^2 \right]^{3/2}} \quad \text{for } i=1,2,\dots,N, \quad (12)$$

Where  $\Delta x_i^j = (x_{i+1}^j - x_{i-1}^j)/2$ , and  
 $\Delta y_i^j = (y_{i+1}^j - y_{i-1}^j)/2$  (13)

$\Delta^2 x_i^j = (\Delta x_{i+1}^j - \Delta x_{i-1}^j)/2$ , and  
 $\Delta^2 y_i^j = (\Delta y_{i+1}^j - \Delta y_{i-1}^j)/2$ . (14)

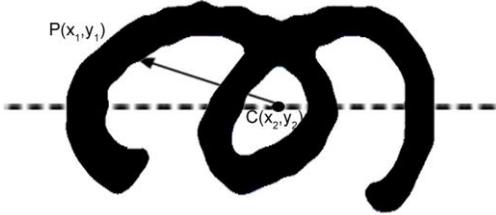


Fig. 8. Distance feature of the ‘tha’ character in Malayalam documents.

C. Distance features

**Distance from the centroid (f8):**

It is distance between the points of the stroke and centroid. Say point P<sub>1</sub>(x<sub>1</sub>, y<sub>1</sub>) on the stroke and centroid has the value of C(x<sub>2</sub>, y<sub>2</sub>) as shown Fig.8.

Then distance between the points P<sub>1</sub> and C is

$$Distance = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2} \quad (15)$$

V. IMPLEMENTATION

Absence of dataset of handwritten pages of different writers in Malayalam posed a great challenge in implementing the scheme. Hence, we collected the handwritings of 280 different users of similar as well as variable content of Malayalam text. The images have been scanned at 400 dpi, 8 bits/pixels, gray-scale. A total of 280 writers contributed to the data set with 50 writers having only one page, 215 writers with at least 2 and 15 writers with at least 4 pages. Each page consists of 21 lines of words with a minimum 30 characters in each line. We kept only the first two images for the writers having more than two pages and divided the image into two parts for writers who contributed a single page thus ensuring two images per writer, one used in training while the other in testing. It has not been possible to test our system against other scripts due to non-availability of the datasets in other scripts. However, a writer identification system for the Indic script, Bengali, was developed by Garain and Paquet [54], which was evaluated on a test bed of only 40 writers, resulting in 75% accuracy.

If the images were are all scanned at the same resolution the size of characters in the writing is a writer-dependent attribute and should not be normalized. For writing instrument independency, the distribution of ink widths in the validation data set were examined and normalized each writing to a fixed

thickness using a thinning algorithm. This, however, resulted in a degradation of performance as some writer-specific information is also lost during the normalization. As of now we decided not to normalize the ink thickness.

Prior to the extraction of features of each writer, the naïve image of documents should be pre-processed at first hand through the following steps.

1. Calculate the threshold value by selecting the image mean and the standard deviation and normalize.
2. Convert the grey-scale of the documents to binary image using the above threshold.
3. Image de-noising is practiced to attain the perfect binary image of the documents [55] [56].

The preprocessed document is then segmented into words using RLSA [57] and Recursive XY Cuts [58] methods. Words are split into characters using the standard connected components algorithm [59]. This connected component analysis is straight-forward, efficient and easy to implement for the segmentation of each character. The knowledge feature vector is created then using the feature extraction method as described in the Section IV. This feature vector is used for the training as well as for the identification phase. For the purpose of classification, K-NN classifier is used for both training and identification phase. For a questioned document the performance is calculated only for the nearest neighbor (top 1) and thus aids in identifying the correct writer.

VI. EXPERIMENTAL RESULTS

The curvature feature was calculated using the two methods as described in Section IV and their efficiency is measured with respect to the number of writers and the observations are given in Table I. The result leads to the conclusion that the contour based curvature feature is well suitable for lengthy documents of large number of writers.

TABLE I. COMPARISON OF POINT BASED AND CONTOUR BASED CURVATURE FEATURE

Number of writers	Point based curvature feature (%)	Contour based curvature feature (%)
10	100	100
25	100	100
50	98	100
75	96	100
100	90	98
150	84.66	97
200	82.5	96.5
250	80.8	95.6
280	78.571	93.92

Usually the slant angle (f6) between each writer makes prominent results in the writer identification but when considering the Malayalam scripts, this variation is very low. Henceforth to identify the minute slant variation between the characters of each writer, f6 is computed using the two

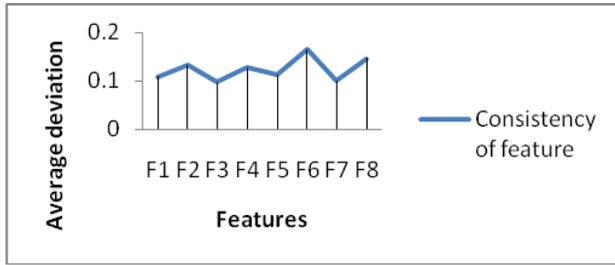


Fig. 9. Consistency of features.

different methods as described in Section IV. It was also observed that the slant angle (f6) calculation based on the histogram outperforms the slant calculation based on the structural features by maxima and minima of the character. The consistency of all the features described in Section IV (f1-f8) is calculated using the mean of standard deviation corresponding stochastic value of each feature and is depicted by the graph given in Fig.9. It is evident that the slant of the character (f6) is relatively a inconsistent feature in the writer identification of Malayalam documents.

Also the experiments yielded us with the performance of each feature in the system of writer identification of Malayalam documents as shown in the following Table. II (numbers represent percentage). The recognition rate is calculated as

$$\text{Recognition rate (Accuracy)} = \frac{\text{Number of correctly identified writer}}{\text{Total number of writers}} \quad (16)$$

The wavelet transform have good spatial frequency localization property which can preserve spatial and gradient information of handwriting [60]. However, to compare our system with existing ones, a Wavelet Domain Local Binary Pattern (WD-LBP) [61] based writer identifications scheme is implemented and the recognition rate is presented in Table III. It is observed that the system achieved a recognition rate of 70.35%.

## VII. CONCLUSION

In this paper, a writer identification scheme for Malayalam language is proposed and implemented. The challenges posed by the scheme resulted in identifying the feature vector for Malayalam script exclusively. The feature vector for Malayalam script has been chosen here with great effort, taking into consideration the special characteristics of the language. The feature vector obtained here is a fusion of eight features which is further used in the training and identification phase. However, the consistencies of the features were studied in detail, and it has been observed that the slant angle feature (f6), which exhibits more prominence for almost all languages, exhibited inconsistency behavior with Malayalam documents. The entire scheme was tested on handwritten documents of 280 writers and obtained a success rate of 95.92%.

TABLE II. COMPARATIVE EVALUATION OF FEATURES

Number of writers	f1	f2	f3	f4	f5	f6	f7	f8	Combined feature (f1+f2+f3+f4+f5+f6+f7+f8)
10	80	70	90	80	80	60	100	70	100
25	80	68	88	80	80	56	100	64	100
50	78	66	88	78	80	56	100	64	100
75	77.33	65.33	88	78	78.66	54.66	100	64	100
100	77	63	87	76	77	54	98	62	99
150	76	62	86.66	74.66	75.33	52.66	97	60.66	99.05
200	75	62	86	74	75	51	96.5	59	98.5
250	74	61.2	84.8	72.8	73.6	49.2	95.6	57.2	97.75
280	72.85	59.64	82.5	70.71	71.78	47.5	93.92	55.35	95.92

TABLE III. PERFORMANCE BASED WD-LBP FEATURE

Number of writers	WD-LBP
10	90%
25	88%
50	86%
75	85.33%
100	83%
150	78%
200	73.5%
250	71.6%
280	70.35%

## REFERENCES

- [1] S. Srihari, S. Cha, H. Arora, and S. Lee, "Individuality of Handwriting," *J. Forensic Sciences*, vol. 47, no. 4, July 2002, pp. 1-17.
- [2] R.C. Gonzalez, R.E. Woods and S.L. Eddins, "Digital Image Processing Using Matlab, Reading", MA: Addison-Wesley, 2004.
- [3] R. Plamondon and G. Lorette, "Automatic Signature Verification and Writer Identification—The State of the Art," *Pattern Recognition*, vol. 22, no. 2, 1989, pp. 107-131.
- [4] F. Leclerc and R. Plamondon, "Automatic signature verification: The state of the art—1989–1993," *Int. J. Patt. Recognit. And Artificial Intell.*, vol. 8, no. 3, June 1994, pp. 643–660..

- [5] A. Schlapbach, L. Marcus, H. Bunke, "A writer identification system for on-line whiteboard data", *Pattern Recognition Journal* 41 (2008) 23821–23897.
- [6] L. Schomaker, "Advances in Writer identification and verification", in: *Ninth International Conference on Document Analysis and Recognition (ICDAR)*, 2007.
- [7] S. Srihari, M. Beal, K. Bandi, V. Shah, and P. Krishnamurthy, "A Statistical Model for Writer Verification," *Proc. Eighth Int'l Conf. Document Analysis and Recognition (ICDAR)*, 2005, pp. 1105-1109.
- [8] S. N. Srihari, S.-H. Cha, and S. Lee, "Establishing Handwriting Individuality Using Pattern Recognition Techniques," in *Proceedings of the Sixth International Conference on Document Analysis and Recognition*, 2001, pp. 1195-1204.
- [9] E. Zois and V. Anastassopoulos, "Morphological Waveform Coding for Writer Identification", *Pattern Recognition*, vol. 33, no. 3, Mar. 2000, pp. 385-398.
- [10] B. Zhang, S. Srihari, and S. Lee, "Individuality of Handwritten Characters," *Proc. Seventh Int'l Conf. Document Analysis and Recognition (ICDAR)*, 2003, pp. 1086-1090.
- [11] S. Al-Ma'adeed, E. Mohammed, D. Alkassis, F. Al-Muslih, "Writer identification using edge-based directional probability distribution features for Arabic words", in: *IEEE/ACS International Conference on Computer Systems and Applications (AICCSA)*, 2008, 582--590.
- [12] H. Said, T. Tan, and K. Baker, "Personal Identification Based on Handwriting," *Pattern Recognition*, vol. 33, no. 1, 2000, pp. 149-160.
- [13] H. Said, G. Peake, T. Tan, and K. Baker, "Writer Identification from Non-Uniformly Skewed Handwriting Images," *Proc. Ninth British Machine Vision Conf.*, 1998, pp. 478-487.
- [14] T. Tan, "Rotation Invariant Texture Features and Their Use in Automatic Script Identification," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 20, no. 7, July 1998, pp. 751-756.
- [15] Y. Zhu, T. Tan, and Y. Wang, "Font Recognition Based on Global Texture Analysis," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 10, Oct. 2001, pp. 1192-1200.
- [16] A. Bensefia, T. Paquet, and L. Heutte, "A Writer Identification and Verification System," *Pattern Recognition Letters*, vol. 26, no. 10, Oct. 2005, pp. 2080-2092.
- [17] A. Bensefia, T. Paquet, and L. Heutte, "Handwritten Document Analysis for Automatic Writer Recognition," *Electronic Letters on Computer Vision and Image Analysis*, vol. 5, no. 2, Aug. 2005, pp. 72-86.
- [18] A. Bensefia, A. Nosary, T. Paquet, and L. Heutte, "Writer Identification by Writer's Invariants," *Proc. Eighth Int'l Workshop Frontiers in Handwriting Recognition*, Aug. 2002, pp. 274-279.
- [19] A. Bensefia, T. Paquet, and L. Heutte, "Information Retrieval Based Writer Identification," *Proc. Seventh Int'l Conf. Document Analysis and Recognition (ICDAR)*, Aug. 2003, pp. 946-950.
- [20] M. Bulacu, L. Schomaker, A. Brink, "Text-independent writer identification and verification on offline Arabic handwriting, in: *Ninth Conference on Document Analysis and Recognition (ICDAR)*, 2007.
- [21] M. Bulacu, L. Schomaker, "Text-independent writer identification and verification using textural and allographic features", *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)* 29(4)(2007)701–717 Special Issue—Biometrics: Progress and Directions.
- [22] H. Zhenyu, Y. Xinge, Y. Y. Tang, "Writer Identification using global wavelet-based features", *Neurocomputing* 71(2008), 1832–1841.
- [23] S. Srihari, C. Tomai, B. Zhang, and S. Lee, "Individuality of Numerals," *Proc. Seventh Int'l Conf. Document Analysis and Recognition (ICDAR)*, 2003, pp. 1096-1100.
- [24] B. Zhang and S. Srihari, "Analysis of Handwritten Individuality Using Word Features," *Proc. Seventh Int'l Conf. Document Analysis and Recognition (ICDAR)*, 2003, pp. 1142-1146.
- [25] C. Tomai, B. Zhang, and S. Srihari, "Discriminatory Power of Handwritten Words for Writer Recognition," *Proc. 17th Int'l Conf. Pattern Recognition*, 2004, pp. 638-641.
- [26] U.-V. Marti, R. Messerli, and H. Bunke, "Writer Identification Using Text Line Based Features," *Proc. Sixth Int'l Conf. Document Analysis and Recognition (ICDAR)*, Sept. 2001, pp. 101-105.
- [27] C. Hertel and H. Bunke, "A Set of Novel Features for Writer Identification," *Proc. Fourth Int'l Conf. Audio and Video-Based Biometric Person Authentication*, 2003, pp. 679-687.
- [28] V. Pervouchine, G. Leedham, "Extraction and analysis of forensic document examiner features used for writer identification, *Pattern Recognition Journal* 40 (2007)1004–1013.
- [29] A. Schlapbach, H. Bunke, "A writer identification and verification system using HMM based recognizers", *Pattern Analysis Application (Springer)*, 10 (2007), 33–43.
- [30] A. Schlapbach, H. Bunke, "Writer identification using an HMM-based hand- writing recognition system: to normalize the input or not?", in: *12th Conference of the International Graphonomics Society, Salerno, Italy, June 26–29, 2005*, pp.138–142
- [31] G. Leeham, S. Chachra, "Writer identification using innovative binerized features of handwriting numerals", in: *Seventh International Conference on Document Analysis and Recognition (ICDAR)*, 2003.
- [32] M. Bulacu, L. Schomaker, L. Vuurpijl, "Writer identification using edge-based directional features", in: *Seventh International Conference on Document Analysis and Recognition (ICDAR)*, 2003
- [33] B. Helli, M.E. Moghaddam, "A text-independent Persian writer identification system using LCS based classifier", in: *IEEE International Symposium on Signal Processing and Information Technology, ISSPIT 2008*.
- [34] S. S. Ram, M. E. Moghaddam, "Text-independent Persian writer identification using fuzzy clustering approach", in: *International Conference on Information Management and Engineering (ICIME)*, Malaysia, 2009.
- [35] M. Soleymani Baghshah, S. Bagheri Shouraki, S. Kasaei, "A novel fuzzy classifier using fuzzy LVQ to recognize online Persian handwriting", in: *Second IEEE Conference on Information and Communication Technology (ICTTA)*, 2006.
- [36] M. Bulacu and L. Schomaker, "Writer Style from Oriented Edge Fragments," *Proc. 10th Int'l Conf. Computer Analysis of Images and Patterns*, Aug. 2003, pp. 460-469.
- [37] G. X. Tan, C. Viard-Gaudin, and A. Kot, "Automatic Writer Identification Framework for Online Handwritten Documents Using Character Prototypes," *Pattern Recogn.*, 2009.
- [38] V. Pervouchine and G. Leedham, "Extraction and analysis of document examiner features from vector skeletons of grapheme 'th'," in *Document Analysis Systems*, pp. 196–207, 2006.
- [39] X. D. Xianliang Wang and H. Liu, "Writer identification using directional element features and linear transform," in *International Conference on Document Analysis and Recognition*, 2003
- [40] T. N. Tan, "Texture feature extraction via visual cortical channel modeling", in *International Conference of Pattern Recognition*, vol. 3, 1992, pp. 607–610.
- [41] T. Y. Zhenyu, He and Y. Xinge, "A contourlet-based method for writer identification," *International Conference on Systems, Man and Cybernetics*, vol. 1, October 10-12 2005, pp. 364–368.
- [42] J. Y. Z. He, B. Fang and X. You, "A novel method for off-line handwriting-based writer identification," in *International Conference on Document Analysis and Recognition*, 2005, pp. 242–256.
- [43] L. Schomaker and M. Bulacu, "Automatic writer identification using connected component contours and edge-based features of uppercase western script", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, June 2004, pp. 787–798.
- [44] A. Seropian, M. Grimaldi, and N. Vincent, "Writer identification based on the fractal construction of a reference base", in *International Conference of Document Analysis and Recognition*, 2003 pp. 1163–1167.
- [45] B. Arazi, "Handwriting identification by means of run-length measurements", *IEEE Transactions of System, Man and Cybernetics*, vol. 7, 1977, pp. 878–881.
- [46] B. Arazi, "Automatic handwriting identification based on the external properties of the samples," *IEEE Transactions of System, Man and Cybernetics*, vol. 13, 1983, pp. 635–642.
- [47] K. Zimmerman, M. Varady, "Handwriter identification from one-bit quantized pressure patterns", *Pattern Recognition*, vol. 18, no. 1, 1985 pp. 63–72.
- [48] Y. Solihin, "A toolset of image processing algorithms for forensic document examination. Technical Report", School of Applied Science, Nanyang Technological University, Singapore, 1997.
- [49] P. Sutanto, G. Leedham, V. Pervouchine, "Study of the Consistency of some Discriminatory Features used by Document Examiners in the Analysis of Handwritten letter 'a' ". In: *the seventh International Conference on Document Analysis and Recognition*, 2003.

- [50] Wei Deng; Qinghu Chen; Yucheng Yan; Chunxiao Wan; , "Off- Line Chinese Writer Identification Based on Character-Level Decision Combination," Information Processing (ISIP), 2008 International Symposiums on , vol., no., 23-25 May 2008, pp.762-765.
- [51] Mandun Zhang; Na Lu; Ming Yu; Xuefeng Zhou; , "The Writer Identification Algorithm Based on Subspace," Intelligent Networks and Intelligent Systems, 2009. ICINIS '09. Second International Conference on Intelligent Networks and Intelligent Systems , 1-3 Nov. 2009
- [52] A. Rehman, D. Mohammad, G. Sulong, "Simple and Effective Technques for Core-region Detection and Slant Correction in Script Recognition" Proceedings of IEEE, International Conference on Signal and Image Processing, 2009.
- [53] E Kavallieratou, N Fakotakis, G Kokkinakis, "Slant estimation algorithm for OCR systems", Pattern Recognition, Vol.34, no.12, Dec.2001, pp.2515-2522.
- [54] U. Garain and T. Paquet, "Off-Line Multi-Script Writer Identification Using AR Coefficients", In Proc.10<sup>th</sup> International Conf. on Document Analysis and Recognition, 2009, pp. 991-995.
- [55] A.Al-Dmour,R.A.Zitar, "Arabic writer identification based on hybrid spectral-statistical measures", Journal of Experimental & Theoretical Artificial Intelligence 19(4)(2007)307–332.
- [56] G.S. Peake and T.N. Tan, "Script and language identification from document images", in Proc. of the British Machine Vision Conference (BMVC97), 1997, Vol. 2, pp. 169–184.
- [57] K. Wong, R. Casey, and F. Wahl, "Document analysis system". *IBM J. Research and Development*, 26(6),1982.
- [58] G. Nagy, S. Seth and M. Viswanathan, "A prototype document image analysis system for technical journals", *IEEE Comput.* **25** (July 1992), pp. 10–22.
- [59] Gonzalez, R.C., Woods, R.E., Eddins, S.L.: Digital Image Processing Using Matlab, Reading, MA: Addison-Wesley, 2004.
- [60] Da-Yuan Xu; Zhao-Wei Shang; Yuan-Yan Tang; Bin Fang; , "Handwriting-based writer identification with complex wavelet transform," Wavelet Analysis and Pattern Recognition, 2008. ICWAPR '08. International Conference on , vol.2, 30-31 Aug. 2008, pp.597-601
- [61] Liang Du; Xinge You; Huihui Xu; Zhifan Gao; Yuanyan Tang; , "Wavelet Domain Local Binary Pattern Features For Writer Identification," 20th International Conference on Pattern Recognition (ICPR), 2010, v23-26 Aug. 2010, pp.3691-3694.