Chapter 3

TEXT INDEPENDENT WRITER IDENTIFICATION

This chapter presents two algorithms to address the problem of text independent writer identification. The database of Roman, Devanagari and Kannada handwritten documents are collected from 100 writers for the proposed writer identification research, because there is no well-known database of Roman, Devanagari and Kannada handwritten documents (3 scripts documents) all written by same writers available for writer identification problem. The documents of IAM [107] English writers is also used as a standard dataset. The results are obtained for 100 writers of single script at a time in Roman, Devanagari, Kannada scripts and also 100 writers in Roman from the standard IAM dataset. The different sized (whole page, half page and quarter page) handwritten documents are used for writer identification to see the potentiality of the features.

Part of this chapter is published in:

3.1 Introduction

The author identification of a handwritten sample document with automatic image-based methods is an interesting pattern recognition problem with direct applicability in the field of forensic and historic document analysis, verification of claimed security, countering repudiation of handwritten mails, determining authenticity of bank transactions, criminal justice systems and archiving documents in library and many more.

As the text-independent methods use any text to establish the identity of the writer; hence, they pose less constraints and are more suitable for real applications. The major challenges in any writer identification system arises from the variability in style, shape, size and consistency of the allographs written by same person at different times and places. In this chapter, two methods are proposed for text independent approach of writer identification.

Recently, a number of new approaches to writer identification have been proposed in the literature. The individuality of handwriting by extracting a set of macro (global) and micro (local) features is established in [86]. Psychological research has shown evidence that the human brain does a frequency analysis of the image [35]. Gabor filters have been found to be good model of the processing that takes place in the human visual cortex, and have been used successfully in both texture segmentation [105] and texture classification [105]. The text area contains high frequency components. This property has been exploited while taking Gabor filter bank approach. These filters are designed to take care of this property [76, 77]. In principle any texture analysis technique [104] can be applied to the uniform text blocks. Texture features based on the co-occurrence histograms of wavelet decomposed images are extracted for off-line text-independent writer identification based on English and Kannada handwriting in [44]. The textures (Kannada script images) are decomposed using Empirical mode decomposition, which in turn generates series of intrinsic mode functions for writer identification of handwritten Kannada script is shown in [57]. Jayanthi et al. [88] proposed text independent method based on the
features extracted from gray level co-occurrence matrices of the scanned image blocks of Tamil handwritten documents of 70 writers. They have used five distances $d=1, 2, 3, 4, 5$ and four directions $\theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ$ to generate a set of 20 co-occurrence matrices. For each of the Gray Level Co-occurrence Matrices they computed the four properties Contrast, Correlation, Energy and Homogeneity, and obtained a total of 80 features for each input block image. The feature vector of the test image is correlated with the feature vectors of train images and the maximum correlation-coefficient was taken as the indicator of the writer of the corresponding test image and obtained the identification rate of 82.8%. In [23] the problem of writer verification is addressed by casting it as a two class problem, authorship and non-authorship. Hanusaik et al. [79] proposed texture based (Gray Level Co-occurrence Matrix) features for writer verification. Textures of the handwritings are created based on the inherent properties of the writer. Independent of the writing style, the proposed method reduces the spaces between lines, words, and characters, producing a texture that keeps the main features thus avoiding the complexity of segmentation. Five properties namely entropy, homogeneity, dissimilarity, inverse variance and energy are extracted from image blocks for the combination of five distances ($d = 1, 2, 3, 4, 5$) and four angles ($\theta = 0, 45, 90, 135$). Franke et al. [36] proposed Gray Level Co-occurrence Matrix based features for ink texture analysis of writer identification. Feature selection was done using Fischer Discriminate Analysis. Ink type inferred from the stroke texture was further used in writer identification.

Golnaz et al. [40] obtained a wide variety of fragments in the connected components from Farsi handwriting. Feature vector was created using the occurrence histogram of the fragments in codebook of the handwriting. A method was proposed to extract fragments and compute the feature vector. To evaluate the method, writer identification was conducted on three varieties of a Farsi database. These varieties include texts of short, medium and large lengths written in 3, 2 and 1 A4 pages respectively from 180 persons. Their experimental results have shown 92.7%, 99.4% and 100% writer identification rate for short, medium and large texts. Biswas et al. [91] presented an approach for extracting two different sets of components (essentially fragments of characters); namely fragment set-A and fragment set-B.
Each extracted fragment was resized to $60 \times 40$. For each resized fragment Radon transform projection profile for the orientations $0^0, 45^0, 90^0$ and $135^0$ was computed. These features from each element of the two sets were used to identify the writing style of a particular person. Using Besus database of Bangla writings they could achieve identification accuracy of $92.72\%$ and $80\%$ for $90$-$110$ words and $20$-$30$ words respectively i.e. by using lesser amount of information from the handwritten samples. Sreeraj et al. [98] used only the character level features like loop, directional and distance features because majority of characters in Malayalam contain loops. However the characters which don't form a loop cannot be so decisive in the feature vector. Combination of all features were tested on handwritten documents of 280 writers and they have achieved success rate of $95.92\%$ using kNN classifier and $70.35\%$ using performance based on WD-LBP method. Sukalpa Chanda et al. [102] collected 2 sets of sample block of 50-60 Oriya words per writer from 100 writers. Directional chain-code and curvature based features were extracted and support vector machine was used for writer identification and obtained $94.00\%$ accuracy. Siddiqi et al. [54] extracted a set of features from contours of handwritten images at different observation levels. At the global level, they extracted the histograms of the chain code, the first and second order differential chain codes and, the histogram of the curvature indices at each point of the contour of handwriting. At the local level, the handwritten text was divided into a large number of small adaptive windows and within each window the contribution of each of the eight directions (and their differentials) was counted in the corresponding histograms. Two writings were then compared by computing the distances between their respective histograms. Using IAM database 250 writers with 2 samples per writer, achieved an identification rate of $86\%$ using chi-square distance with minimum distance (kNN with $k=1$) classifier.

Jing et al. [55] used multiresolution approach to off-line, text-independent Chinese writer identification based on edge structure code (ESC) distribution feature and nonparametric discrimination of sample. ESC distribution feature is based on probability distribution function, which characterized the frequent structures distribution of edge fragments on multiple scales. From an open HIT-MW database, 240 text documents were used which were provided by 240 writers and split each of
them into two sub images, with one training handwriting and one query handwriting and evaluated the performance of 95.4% in writer identification. Schomaker et al. [59] used features of the contours of fragmented connected components in mixed-style handwritten samples of limited size. The writer was considered to be characterized by a stochastic pattern generator, producing a family of character fragments (fraglets). Using a codebook of such fraglets from an independent training set, the probability distribution of fraglet contours was computed for an independent test set. Each of the 150 paragraphs of the 150 writers was divided into a top half (set A) and a bottom half (set B). Writer descriptors p(FCO3) were computed for set A and B. Using Kohonen self-organized map (KSOM) of fragmented connected component contours of free style writing achieved a performance of 70% which might be attributed to both the smaller number of characters and its variable text content. Hertel et al. [22] used a subset of IAM database i.e. of 250 pages written by 50 writers. Each writer contributed 5 pages of text. One page comprised about 8 lines of text on the average. The total data set consists of 2185 lines of text. Features extracted were connected components, enclosed regions, lower and upper contour, fractal features, basic features, all features of single line and also combination of all lines. Using majority voting, 5-NN classifier they achieved 90.7% writer identification accuracy with all features of single line from each writer and also shown that by the combination of all lines of each document the writer identification rate increases to 99.6%.

3.2 Dataset

A sample of 100 writers is chosen from students of different ages between 17-25 for identification who know to write in all the 3 scripts (Roman, Devanagari and Kannada). The writers are not imposed any constraint like type of pen, style of writing etc., and the purpose of data collection is also not disclosed. To acquire the handwriting from persons, the writers are provided with the A4 size unrolled papers and are informed to write any text matter (may be from magazines, newspapers or on their own) of four pages each in Roman (English), Devanagari (Hindi) and Kannada scripts. A total of 1200 document images are created from 100 writers (4
pages X 3 scripts X 100writers=1200 documents). Each document is written in single script only. The collected documents are scanned using HP scanner at 300 DPI for converting them into digital documents, which usually yields a low noise and good quality document images of size 3507X2250. The four pages of texture (3507X2250 pixels) images from each writer, this adds up to 400 texture images in each script. Part of IAM dataset [107] of 4 pages each of 100 writers is used as a standard dataset for the proposed text independent writer identification.

The digitized images are converted into binary images using Otsu's global thresholding method. To preserve writer specific details of directional strokes the skew correction of document, slant correction of the lines of document are not applied. A sample page of English handwritten document image of a writer from created dataset is presented in Figure 3.1, two sample documents of IAM handwritten dataset are presented in Figure 3.2.

Figure 3.1: Sample page of English handwritten document of a writer from the created dataset
Chapter 3: Text Independent Writer Identification

He, Kennedy told Mr. Macmillan that he still wanted him to apply for membership of the Common Market, even if it meant an unconditional surrender.

There were also brief discussions on East Berlin and other foreign questions, after Mr. Kennedy had informed Mr. Macmillan of his discussions with Mr. Kruzhkov. With the exception of 40 minutes when Lord Home, Foreign Secretary, and Mr. McGeorge Bundy, the President's special assistant for security affairs, were brought in, the two men had talked alone.

But there is heart in the telling, and an incipient warmth in the situation. A young girl lives in a simple drab room with her mother, prominent mother. In much surrounding she seems more than something normal; and when she experiences it for the first time herself it is incomparably charming, and half simply and half ingenuously. As in the case of Terri, this girl lives in a studio where leaves has to bear her child and raise away.

**Figure 3.2:** A couple of sample pages of two different writers from IAM Handwriting dataset

The schematic overview of writer identification is presented in Figure 3.3.

**Figure 3.3:** Schematic overview of writer identification
3.3 Writer Identification

Writer identification consists of several steps: pre-processing, feature extraction, similarity measurement, performance evaluation. Among these steps, the feature extraction is the core one.

As each writer has unique characteristics of horizontal, vertical and slanted lines (strokes) as well as spatial characteristics in his writings, two methods viz directional stroke feature based method and gray level co-occurrence matrix feature based method have been devised for text independent writer identification. The first method exploits the eight directional stroke based features of the handwritten document image. The second method exploits the different texture descriptors extracted from Gray level co-occurrence matrices along 4 directions and 5 distances of the handwritten document image.

3.3.1 Directional Stroke Features

According to the experts, a given writer takes some directions more frequently and more intensively than others in his writings. Therefore directional stroke based features are discriminative features of a writer in the proposed method. The important characteristic of the shape of Roman script is combination of vertical strokes, left or right slanted lines (slanted strokes) with few horizontal strokes in its characters, of Devanagari script is combination of horizontal, vertical, left slanted and right slanted strokes in its characters, of Kannada script is combination of horizontal and vertical strokes, and has more number of circular shapes. The directional strokes of different angles play a significant role in distinguishing scripts. These directional stroke features are extracted from the connected components of an image. The writers of documents have their own style of writing, some writers may use large strokes or small strokes or right slanted strokes or left slanted strokes of different angles and different lengths for the given sample text, i.e., each writer has his own specific characteristics of strokes in his writings which are represented in the form of different oriented directional strokes of different lengths for the given sample text. This motivates to consider the directional strokes as the feature vector for writer identification.
3.3.1.1 Methodology and Algorithm

The proposed idea is inspired by the work of Dhandra et. al [11]. Because every writer has uniqueness in structure of strokes of his writings, in this chapter a novel method is proposed for exploiting directional stroke based features of the connected components of input binary document image at page level for writer identification.

Before feature extraction, using connected component analysis, the components of area less than 50 were removed which include small objects like commas, single or double quotes which do not contribute much to writer or script identification. The erosion operation followed by morphological opening is performed on the connected components of input binary image in eight directions horizontal, vertical, right and left diagonal directions and at angles $0^\circ$, $30^\circ$, $45^\circ$, $60^\circ$, $90^\circ$, $120^\circ$, $135^\circ$ and $150^\circ$ with the line-structuring element to obtain the strokes present in the connected components of an image. The length of the structuring element 7 is fixed through experiment. For a given stroke the definition of length, width and height of the stroke in the proposed work are illustrated in Figure 3.4.

![Figure 3.4: Meaning of length, width and height of a stroke in the proposed method.](image)

The input handwritten document image, pre-processed images, its connected components and output images of erosion followed by opening of the connected components along eight directions are presented in Figure 3.5.
Chapter 3: Text Independent Writer Identification

Figure 3.5: A sample (a) Input colored image block   (b) Binary image of (a) (c) Complimented image of (b) (d) Border Cleared Image of (c) (e) Preprocessed image of (d) (f1), (f2), (f3) and (f4) are Connected Components of (e) (g1), (g2), (g3) and (g4) are output images of erosion followed by opening of (f1), (f2), (f3) and (f4) at angle 0° (h1), (h2), (h3) and (h4) are output images of erosion followed by opening of (f1), (f2), (f3) and (f4) at angle 30° (i1), (i2), (i3) and (i4) are output images of erosion followed by opening of (f1), (f2), (f3) and (f4) at angle 45° (j1), (j2), (j3) and (j4) are output images of erosion followed by opening of (f1), (f2), (f3) and (f4) at angle 60° (k1), (k2), (k3) and (k4) are output images of erosion followed by opening of (f1), (f2), (f3) and (f4) at angle 90° (l1), (l2), (l3) and (l4) are output images of erosion followed by opening of (f1), (f2), (f3) and (f4) at angle 120° (m1), (m2), (m3) and (m4) are output images of erosion followed by opening of (f1), (f2), (f3) and (f4) at angle 135° (n1), (n2), (n3) and (n4) are output images of erosion followed by opening of (f1), (f2), (f3) and (f4) at angle 150°

The writer specific stroke details are captured using directional stroke based features, because average stroke length and number of strokes per unit length of connected components varies from writer to writer.
Even though this work is motivated by Dhandra et al. [11], the features like Stroke Length and Stroke density used in this chapter are different. They defined stroke length as the number of pixels in a stroke as the measure of its length, for the strokes in vertical, horizontal, right and left diagonal directions of the image. Further, they defined the stroke density as the total length of all the strokes divided by the size of the image.

In the proposed method, thinning operation is not applied for the input document image in order to consider the thin lined (faintly visible) characters of the handwritten documents of some writers. For writers having thick handwritings, if the number of pixels in a stroke is considered as a measure of its length and it will be more than its actual length. Because the writer may write using different pens at different times with different pressures, the thickness of strokes vary, to make the writer or script identification independent of the thickness of handwritten strokes, the definition of stroke length and stroke density have to be modified and are presented in the following:

**Stroke Length (SL)**

It is Euclidean distance between two end pixels of the stroke as the measure of its length. Average stroke length is average of the average stroke length of the individual strokes found in the connected component of an image. \( N \) is the number of connected components found in an image and \( n \) is number of strokes found in a \( j^{th} \) connected component of the image.

Compute the average stroke lengths in 8 directions at angles \( \theta = 0^0, 30^0, 45^0, 60^0, 90^0, 120^0, 135^0 \) and \( 150^0 \) and are computed using the following equations:

**Average Stroke Length (ASL\(_\theta\))**

\[
\text{ASL}_{\theta}(\text{pattern}) = \frac{1}{N} \sum_{j=1}^{N} \left( \frac{1}{n} \sum_{i=1}^{n} \text{len}(s_i)_{\theta} \right) \\
\]

\( n \) is the number of strokes in the \( j^{th} \) connected component, \( s_i \) is the \( i^{th} \) stroke and \( N \) is the total number of connected components.
Stroke Density (SD)
The number of strokes per unit length (x-axis) of the connected component is stroke density. Compute stroke densities in 8 directions at $0^0$, $30^0$, $45^0$, $60^0$, $90^0$, $120^0$, $135^0$ and $150^0$ angles. Then the average stroke density of document in each direction is obtained by averaging the stroke densities found in the connected components of an image. The computational formula is given below:

Average Stroke Density (ASD):

$$ASD_{\theta}(\text{pattern}) = \frac{1}{N} \sum_{i=1}^{N} \frac{n_i}{w(c_i)}$$  \hspace{1cm} (3.2)

where $N$ is total number of connected components found in the image, $n_i$ is the number of strokes of the direction $\theta$ in the $i^{th}$ connected component and $w(c_i)$ is width of the $i^{th}$ connected component.

The writer identification algorithm for whole page input documents is given below:
Algorithm: Writer Identification based on Directional Stroke Features
Designed algorithm has of two phases Train Phase and Test Phase. In the Train Phase features from sample images marked for training are extracted and stored. In the Test Phase features are extracted from sample images marked for testing and are stored in as test library. Nearest neighbor, kNN and LDA classifiers may be used for classification.

\begin{itemize}
  \item **Input**: Gray scale images of handwritten document page each of size 3507X2550 pixels.
  \item **Output**: Identification of the writer for input document image
  \item **Method**: Directional Stroke Based method
  \item **Size of the Feature vector**: 16
  \item **Size of the Dataset**: 400 pages of each Script (with 4 pages per script per writer from 100 Writers)
\end{itemize}

**Train Phase:**
Begin
\begin{enumerate}
  \item Binarization: Convert gray scale image to binary image using Otsu's method.
  \item Preprocessing: Clear the image border to remove the unwanted objects connected to an image border.
  \item Feature Extraction: For the preprocessed image, extract strokes of each connected component of an image in eight directions at $0^\circ$, $30^\circ$, $45^\circ$, $60^\circ$, $90^\circ$, $120^\circ$, $135^\circ$ and $150^\circ$ using morphological erosion and opening operations in corresponding directions.
  \item Compute the average stroke lengths and densities of the extracted strokes obtained in the above mentioned directions as mentioned in eqns. (3.1) and (3.2). So that 16 features are obtained.
  \item Trained feature Library: Store the computed feature sets with writer specific labels in the train library.
\end{enumerate}
End

**Test Phase:**
Begin
\begin{enumerate}
  \item Compute the feature vector of test image using steps 1 to 4 of above training phase.
  \item Store the computed feature sets with writer specific labels in the test library.
  \item Identify the writer of test image using Nearest Neighbor / kNN / LDA classifiers.
\end{enumerate}
End
Evaluation protocol

Here modified 4-fold cross validation (CV) is used to evaluate the performance of the proposed method. Here data is split into 4 equal sized sub-parts such that each time among 4 sub-parts of each writer, three sub-parts are considered for training and remaining one sub-part is taken for testing, such all combinations of 3:1 of sub-part of each writer are considered for validation and single value result is calculated by averaging all results.

3.3.1.2 Experimental Results

The experiment is conducted on handwritten document images of 100 writers on Roman (R), Kannada (K) and Devanagari (D) scripts, and also same number of IAM writers. The stroke features are extracted from the handwritten documents using Directional Stroke based method. The stroke length and stroke density in 8 directions are extracted as a feature vector of size 16. The results are obtained by using Nearest Neighbor, kNN classifier with Euclidean distance measure and LDA with modified 4-fold cross validation. The experiment is conducted by considering single scripts Roman, Kannada and Devanagari of the 100 writers and the tabulated results are presented in Table 3.1.

From the Table 3.1, it is evident that for whole page documents of Roman, Devanagari and Kannada, the highest writer identification accuracies are 99.75%, 99.75% and 99.74% respectively and that by using IAM dataset of Roman writers is 96% respectively using LDA classifier. The variation in writer identification rate is due to variation of the directional stroke features from one writer to another writer.
Table 3.1. Experimental Results of Text Independent Writer Identification using Directional Stroke Features

<table>
<thead>
<tr>
<th>Scripts</th>
<th>Accuracy in % using Classifiers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nearest Neighbor with Euclidean Distance</td>
</tr>
<tr>
<td>Roman</td>
<td>96.5946</td>
</tr>
<tr>
<td>Roman_IAM</td>
<td>87.25</td>
</tr>
<tr>
<td>Devanagari</td>
<td>93.507</td>
</tr>
<tr>
<td>Kannada</td>
<td>95.6488</td>
</tr>
<tr>
<td>Roman Half Page</td>
<td>75.375</td>
</tr>
<tr>
<td>Roman_IAM Half Page</td>
<td>55.25</td>
</tr>
<tr>
<td>Devanagari Half Page</td>
<td>74.875</td>
</tr>
<tr>
<td>Kannada Half Page</td>
<td>70.75</td>
</tr>
<tr>
<td>Roman Quarter Page</td>
<td><strong>69.6875</strong></td>
</tr>
<tr>
<td>Roman_IAM Quarter Page</td>
<td><strong>55.125</strong></td>
</tr>
<tr>
<td>Devanagari Quarter Page</td>
<td><strong>65.9375</strong></td>
</tr>
<tr>
<td>Kannada Quarter Page</td>
<td><strong>65.9375</strong></td>
</tr>
</tbody>
</table>

Among all the classifiers used, LDA has highest performance in writer identification for whole page and half page input documents. The identification accuracy of IAM writers is relatively lesser than that of Roman, Devanagari and Kannada (our dataset) writers, this is due to the size of images (pages) of IAM dataset are lesser compared to size of images used in our dataset. It is obvious that lesser the size of image, the lesser will be information content in the image and lesser will be the accuracy. For quarter page input documents, NN classifier with Euclidean distance performs better for directional stroke based writer identification.
3.3.2 Gray Level Co-occurrence Matrix Features

The GLCM based features are described in Section 1.5.5. These features are used in designing an algorithm for writer identification to see how best are these features. The writer identification algorithm for whole page input documents is given below:

### 3.3.2.1 Algorithm: Writer Identification using GLCM Features

**Input**: Gray scale images of handwritten document page each of size 3507X2550 pixels.

**Output**: Identification of the writer for input document image

**Method**: GLCM Based method

**Size of the Feature vector**: 40

**Size of the Dataset**: 400 pages of each script (with 4 pages per script per writer from 100 Writers)

**Train Phase:**

Begin

1. **Binarization**: Convert gray scale image to binary image using Otsu's method.

2. **Pre-processing**: Clear the image border to remove the structures connected to image border.

3. **Feature Extraction**: For the preprocessed image, obtain 20 GLCMs for 4 directions 0°, 45°, 90° and 135° for five distances d=1, 2, 3, 4, 5. For each GLCM extract the Correlation co-efficient and Homogeneity, so that 40 features are obtained.

4. **Train Library**: Store the computed feature sets with writer specific labels in the train library.

End

**Test Phase:**

Begin

1. **Compute the feature vector of test image using steps 1 to 3 of above training phase.**

2. **Store the computed feature sets with writer specific labels in the test library.**

3. **Identify the writer of test image using Nearest Neighbor / kNN / LDA classifiers.**

End
3.3.2.2 Experimental Results

The experiment is conducted on handwritten document images of 100 writers on Roman (R), Kannada (K) and Devanagari (D) scripts, and also same number of IAM writers. The texture features are extracted from the handwritten documents using GLCM based method. The results are obtained by using Nearest Neighbor, kNN classifier with Euclidean distance measure and LDA with modified 4-fold cross validation. The experiment is conducted by considering single scripts Roman, Kannada and Devanagari of the 100 writers and the tabulated results are presented in Table 3.2.

<table>
<thead>
<tr>
<th>Scripts</th>
<th>Accuracy in % using Classifiers</th>
<th>LDA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nearest Neighbor with Euclidean Distance</td>
<td>Nearest Neighbor with City block Distance</td>
</tr>
<tr>
<td>Roman</td>
<td>94.5551</td>
<td>95.0919</td>
</tr>
<tr>
<td>Roman_IA_M</td>
<td>87.6681</td>
<td>89.038</td>
</tr>
<tr>
<td>Devanagari</td>
<td>96.5946</td>
<td>95.1644</td>
</tr>
<tr>
<td>Kannada</td>
<td>95.3622</td>
<td>96.2309</td>
</tr>
<tr>
<td>Roman Half Page</td>
<td>81.5000</td>
<td>78.8750</td>
</tr>
<tr>
<td>Roman_IA_M Half Page</td>
<td>78.2500</td>
<td>77.2500</td>
</tr>
<tr>
<td>Devanagari Half Page</td>
<td>84.0000</td>
<td>80.7500</td>
</tr>
<tr>
<td>Kannada Half Page</td>
<td>82.7500</td>
<td>80.3750</td>
</tr>
<tr>
<td>Roman Quarter Page</td>
<td>78.5</td>
<td>75.6875</td>
</tr>
<tr>
<td>Roman_IA_M Quarter Page</td>
<td>78.75</td>
<td>78.3125</td>
</tr>
<tr>
<td>Devanagari Quarter Page</td>
<td>75.1875</td>
<td>72.1875</td>
</tr>
<tr>
<td>Kannada Quarter Page</td>
<td>81.1875</td>
<td>76.8125</td>
</tr>
</tbody>
</table>
For whole page input documents, the highest writer identification accuracy for Roman (English) is 99.7475%, for Devanagari (Hindi) is 99.2266%, for Kannada script is 99.75%, and that by using IAM dataset of Roman writers is 97% using LDA classifier.

Writer identification for single script input documents the LDA is the better classifier for GLCM based writer identification for all whole page, half page and quarter page input documents of single script. All the classifiers have shown good performance, but the LDA classifier’s performance is relatively higher when compared to the other classifiers. The identification accuracy of IAM writers is relatively less than that of Roman, Devanagari and Kannada (our created dataset) writers, again this is due to the size of images (pages) of IAM dataset are lesser compared to image size of our dataset (3507X2250 pixels). Lesser the information content in the image (lesser the image size), the lesser will be the accuracy.

In proposed method the identification accuracy is 87.6681 % and 84.0428% using KNN classifier for k=1 and k=3 respectively. Said et al. [42] obtained 88.0% accuracy of writer identification considering 60 features using KNN classifier. Though the proposed GLCM based writer identification is an existing method but in the proposed method lesser number of features are used i.e. 40 features (Correlation and homogeneity along 4 directions 0\(^0\), 45\(^0\), 90\(^0\) and 135\(^0\) for five distances d=1, 2, 3, 4, 5).

### 3.4 Summary

In this Chapter, text independent writer identification is carried out based on the proposed two feature extraction techniques with NN, kNN (Euclidean and City Block distances) and LDA classifiers. The proposed feature extraction method is well suited for LDA classifier with maximum accuracy of 99.75% for directional stroke based features and 99.2266% for GLCM features for whole page input documents of Devanagari scripts. Both feature extraction techniques have performance accuracy for Roman and Kannada scripts. In view of the results presented in Table 3.1 and Table 3.2, as the size of handwritten document image is
reduced, the directional stroke based method gives much lesser writer identification accuracy compared to GLCM based method.

Writer identification based on Directional Stroke Based Features is a novel approach, considering the individuality of writer specific strokes of different orientation in their writings. This method is robust with respect to type of pen or pen tip thickness.

It is observed from the results of the two proposed methods, the performance of GLCM based features is generally higher than that of the directional stroke based features.

The contribution of the proposed methods is two folded. One it performs writer identification independent of the allograph segmentation and classification. Second, features proposed show significant results and are robust to noise.