

# Writer Verification Using Spatial Domain Features under Different Ink Width Conditions

**Sharada Laxman Kore\***

Bharati Vidyapeeth Deemed University College of Engineering, Pune, India  
sharadakore@gmail.com

**Shaila Dinkar Apte**

Department of Electronics and Telecommunication, Rajarshi Shahu College of Engineering, Pune, India  
sdapte@rediffmail.com

## Abstract

In this paper, we present a comparative study of spatial domain features for writer identification and verification with different ink width conditions. The existing methods give high error rates, when comparing two handwritten images with different pen types. To the best of our knowledge, we are the first to design the feature with different ink width conditions. To address this problem, contour based features were extracted using a chain code method. To improve accuracy at higher levels, we considered histograms of chain code and variance in bins of histogram of chain code as features to discriminate handwriting samples. The system was trained and tested for 1,000 writers with two samples using different writing instruments. The feature performance is tested on our newly created dataset of 4,000 samples. The experimental results show that the histogram of chain code feature is good compared to other methods with false acceptance rate of 11.67%, false rejection rate of 36.70%, average error rates of 24.18%, and average verification accuracy of 75.89% on our new dataset. We also studied the effect of amount of text and dataset size on verification accuracy.

**Category:** Human computing

**Keywords:** Chain code; Differential chain code; Variance; Writer verification

## I. INTRODUCTION

### A. Motivation

Person identification and verification based on handwriting is in demand due to the great importance in criminal justice for handwriting analysis. Writer identification and verification operates in two modes: identification and verification. A writer identification system, applicable to forensic analysis, performs a one-to-many search in a large

database and returns a likely list of candidates. Writer verification involves a one-to-one comparison with a decision whether or not the two samples are written by the same person. Verification tasks are applicable to security systems and access control systems. Choice of mode of operation of writer identification and verification depends on applications.

The traditional method of determining the identity of a writer, in handwritten document, is tedious, time consuming and suffers from human factors/errors. Therefore,

**Open Access** <http://dx.doi.org/10.5626/JCSE.2016.10.2.39>

<http://jcse.kiise.org>

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 02 November 2015; Accepted 14 March 2016

\*Corresponding Author

Stand up, be bold, be strong. Take the is  
on your own shoulders, and know t  
creator of your own destiny. Three things are  
every man great, every nation great: 1) Con  
hance as a

(a)

when things go wrong in our  
difficult situations we tend to  
itself as the problem, but in  
problems we experience com

(a)

difficult situations we tend to regard the  
situation itself as the problem, but in reality w  
problems we experience come from the side of t  
mind. If we were to respond to difficult situatio  
word with a positive or peaceful mind they w  
be problem for us; indeed we may even come to s  
them as challenges or opportunities.

(b)

When things go wrong in our  
encounter difficult situations  
regard the situation itself c  
but in reality whatever pr

(b)

**Fig. 1.** Samples, from one writer, using ball pen and sketch pen (samples are from new dataset). (a) Handwriting of writer ID 129 using ball pen. (b) Handwriting of writer ID 129 using sketch pen.

efforts are made by researchers, to design automatic writer identification and verification systems. Although much research has been done in this area, the existing methods do not give high accuracy under different ink width conditions. The researchers (except Srihari) presented results on IAM dataset [1], which does not include handwritten samples with variable ink widths. The type of writing instrument used is observed to greatly affect the handwriting. The writing instrument, pen pressure and pen control determine the ink distribution in handwritten document image. The handwriting of the same person using different pen types may vary at large scale as shown in Fig. 1.

As shown in Fig. 1, the two samples of the same writer, with different pen types, look significantly different. The distance between the feature vectors of these two samples may be greater than the threshold of writer 129 and the system may falsely reject two samples, written by the same person. The other possibility is that the two handwritings of two different persons may be similar as shown in Fig. 2.

In case of similar handwriting patterns of two different people, distance between the two feature vectors of different writer samples may be less than the threshold of claimed writer and the system may falsely accept as same writer samples. Fig. 1 shows an example of false rejection and Fig. 2 shows an example of false acceptance.

Obviously, when handwriting samples with different ink width conditions are compared, the average of false acceptance and false rejection error rate may be high and average verification accuracy may be low. The experimental results shows that the average error rate is high under dif-

**Fig. 2.** Different writer samples using sketch pen. (a) Handwriting of writer ID 001 using sketch pen. (b) Handwriting of writer ID 007 using sketch pen.

ferent ink width conditions (ballpoint pen vs. sketch pen) compared to same ink width conditions.

In this paper, we addressed the problem of writer verification under different ink width conditions. In real life, there are many applications where we encounter this problem. Past documents have been found with handwritings using different writing instruments. The handwriting varies under different writing conditions such as the writing by pen on paper (answer sheets, diary, lecture notes, ransom notes), by chalk on blackboard, by marker pen on white board/CD (teaching aids), and by lipstick on mirror (threatening messages). In our work, we consider the pen as the writing instrument and paper as the writing surface.

Evidence that the problem is significant and worth solving is mentioned here. In written examinations, the student can change pens while writing on the answer sheet. The handwriting using ink pens is different from the handwriting using ballpoint pens. The different handwriting patterns can be found for the same writer on the same answer sheet. Handwriting fraud cases may occur in written examinations. In fraud cases, it is difficult to authenticate the answer sheet. In addition personal authentication is required in case of doctor medical prescriptions (in case of death/murder/suicide), will deed, or patent conflicts. In such cases, two available handwritten samples may be taken and compared for authentication of a person. In case of access control systems, false acceptance is more dangerous than false rejection. Experimental results have shown that the false acceptance is very low in the case of spatial domain features for writer verification. Therefore, in this paper, the first objective of our work is to test the usefulness of existing spatial domain methods of writer identification and verification under different ink width conditions. The second objective is to design a new spatial domain feature for writer verification which will work with different ink width conditions.

## B. Related Work

Comparative study of existing methods has been presented by Kore and Apte [2]. Broadly, we can classify the techniques of writer identification and verification as transform domain technique and spatial domain technique. The transform domain technique extracts the feature based on texture image. However, the spatial domain technique extracts the feature based on pixel information. In the spatial domain technique, the probability distribution of pixel information is used as a feature for writer identification and verification. We give below a detailed literature review of spatial domain techniques for off-line text-independent, writer identification and verification.

Arazi [3] presented Run-Length method for writer identification. The histograms of number of black and white pixels in the horizontal and vertical directions were used as features for writer identification, capturing information about the intra-word and inter-word spacing in handwritten document images. The histograms have been found to be different for same writer samples with different writing conditions. We have observed in our experiments that the run length features are not useful with different ink width conditions.

Bulacu and his associate [4-6] presented a texture-level and allograph-level approach using edge-based directional probability distribution functions (PDFs) as features for writer identification that is the most widely used feature for writer identification and verification. The experimental results had shown that the joint PDFs of “hinged” edge-angle combinations outperform all the other evaluated features. The accuracy obtained was 89% on IAM-650 dataset. The problem with this technique is the use of edge detection as preprocessing step for writer identification and verification. Ink distribution depends on the type of pen and pen pressure as well as the style of writer. There will be loss of some useful information after edge detection with different ink width conditions. We have done experimentation and observed that the accuracy is decreased with different ink width conditions.

Every writer acts as a stochastic generator of ink-blob shapes or graphemes. Bensefia [7] used graphemes generated by a handwriting segmentation method to encode the individual characteristics of handwriting independent of the text content. Grapheme clustering is used to define a feature space common for all documents in the dataset. Experimental results were reported on three datasets containing 88 writers, 39 writers (historical documents) and 150 writers, with two samples (text blocks) per writer. Writer identification rate around 90% was reported on the different test datasets.

Akhter [8] has recently proposed script independent method based on branchlet distribution features which were extracted at the document level. The pre-processing steps involved were binarization, counter extraction and

skeletonization. Handwriting was taken as n-branching ink fragments, originating from a central point. PDF of this branching structure’s orientation was used as features which overcome allographic constraints. A promising result of 96% was obtained on the IAM database. The features are independent of the handwriting width and the localization problem of branching origin is also resolved. But the process requires 17 hours for codebook generation of 255 writers using a 2.33 GHz CPU and 3.48 GB RAM.

Brink proposed writer identification using directional ink trace width measurements [9]. Siddiqi and Vincent [10] proposed an effective method based on orientation and curvature. The proposed methodology evaluated two different datasets, exhibiting promising results on writer identification and verification. The authors [11] introduced a set of features that were extracted from the contours of handwritten image at different observation levels. At the global level, the features considered were 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. The system provided an accuracy of 79% and 86% on RIMES (650 writers) and IAM (225 writers) database, respectively.

Recently, Huang et al. [12] proposed differential chain codes and grid features for writer identification and verification. The method operated at three stages. Ding et al. [13] proposed a method based on local contour features. The distribution of local contour was extracted from the fragments, which were parts of the contour in sliding windows. In order to reduce the impact of stroke weight, the fragments which do not directly connect the center point were ignored during feature extraction. The edge point distributions of the fragments were counted and normalized into local contour distribution features (LCDF). Ram and Moghaddam [14] presented Persian writer identification using swarm-based feature selection approach. Chaabouni et al. [15] proposed a new method for writer identification based on multi fractal features using combination of on-line and off-line approaches. The tests were performed on the writing of 110 writers from the ADAB (Arabic) database. Abdi and Khemakhem [16] proposed a model based approach to off-line text independent Arabic writer identification and verification.

The survey of the above methods has shown that the researchers did not explicitly address the issue of the writing instrument, which produces ink width variation in handwritten documents. Siddiqi and Vincent [17] divided the handwritten text into connected components and each component was repeatedly thinned to a point where there

was no change in the component area. The ratio of area to number of iterations was called blob fraction and was used as feature for handwriting classification.

Based on survey and our knowledge, we find that we are the first to address the issue of writer verification under different ink width conditions. Kore and Apte [18] tested the usefulness of chain code and differential chain code features on new character level dataset including samples of variable ink widths.

### C. Main Contributions of Our Work

The current work focuses on a comparative study of the result of spatial domain techniques of off-line text-independent writer verification on a large dataset, under different ink width conditions. To the best of our knowledge, we are the first in creating such a large dataset. We designed new features based on chain code, which provides improved results compared to existing chain codes. The main contributions are as follows.

- Tested the result of most commonly used existing methods under different ink width conditions.
- Designed new features using chain code for writer verification.
- Compared performance of spatial domain features.

The remainder of this paper is organized as follows. Section II explains the newly created dataset. The feature extraction methods are presented in Section III. Writer verification is explained in Section IV. Experimental results and discussions are presented in Section V. Finally, we present the conclusions and future scope in Section VI.

## II. DATASET

The existing datasets (IAM-2002 [1], CEDAR-2000 [19], and IBM-2013 [20]) of off-line English handwriting do not include ink width variations. Therefore, a new dataset containing samples under different ink width vari-

ations was created, with Table 1 providing the comparison of our dataset with existing off-line English handwriting datasets.

We have created a new dataset by collecting handwriting samples from school children, undergraduate students, working people of different age groups, gender and places. The subjects were asked to write a given English text of 5 to 6 lines, using ballpoint pen and sketch pen on a piece of A4 size plain paper. The given text was different for different professionals. The dataset includes 4 samples of each 1,000 writers. Each writer is asked to write the two samples using a ballpoint pen and two samples using a sketch pen. The size of dataset is of 4,000 samples. The handwritten pages were scanned using professional HP scanner with 300 dpi resolution and scanned images are stored in computer as an 8-bit color JPG type images.

## III. FEATURE EXTRACTION

This section presents the new feature using chain code method. The most commonly used existing methods for writer identification and verification are explained below.

### A. Entropy

Entropy gives information about the average ink distribution in the handwritten document image [6] and is a statistical measure of randomness that can be used to characterize the texture of the input image. Entropy (E) is defined as in Eq. (1).

$$Entropy = \sum p. * \log(p) \tag{1}$$

Where,  $p$  contains the histogram count of intensity image. The pen control and applied pressure in the handwritten document is different for each individual. The average ink distribution in the document image gives

**Table 1.** Comparison of the new dataset with existing offline English handwriting datasets

Parameter	Our dataset (2015)	Existing datasets		
		IBM (2013)	IAM (2002)	CEDAR (2000)
Ink distribution for same writer samples	Variable	Equal	Equal	Equal
Text for same writer samples	Different	Different	Different	Same
No. of writers	1,000	47	657	900
Samples/writer	4	-	2	3
Total no. of samples	4,000	2,000	1,539	2,700
Resolution (dpi)	300	-	300	300
Image type	JPG 8 bit, color	-	PNG, 8 bit, grayscale	BMP

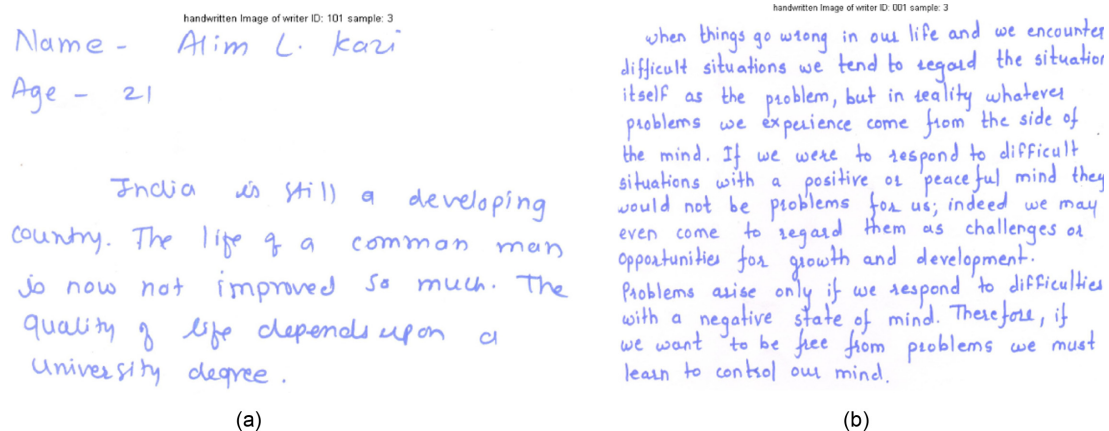


Fig. 3. Entropy of different writer samples. Entropy of images (a)  $E1=2.3199$  and (b)  $E2=2.3467$ .

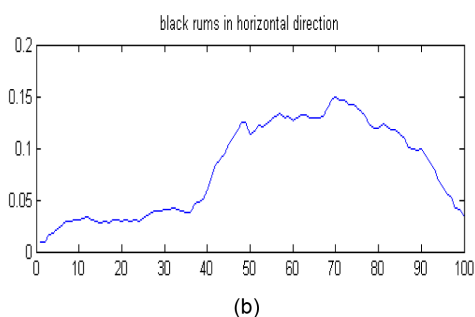
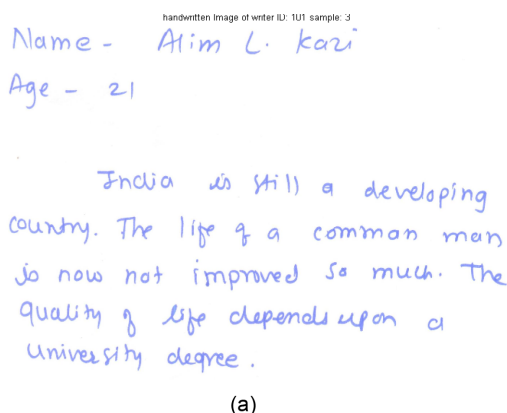


Fig. 4. Run length feature. (a) Original image of writer ID 101 sample 3 and (b) black runs of given image in (a) in horizontal direction.

information regarding the pen pressure of a writer and the type of writing instrument used. Therefore, we used entropy as a feature for writer verification.

First, the original color image is converted into grayscale. Then, entropy is calculated using MATLAB function. Fig. 3 shows entropy for same writer samples under roughly equal ink distribution and unequal ink distribution.

The size of feature vector is single element. Fig. 3 demonstrates that the entropy is different for different writer samples.

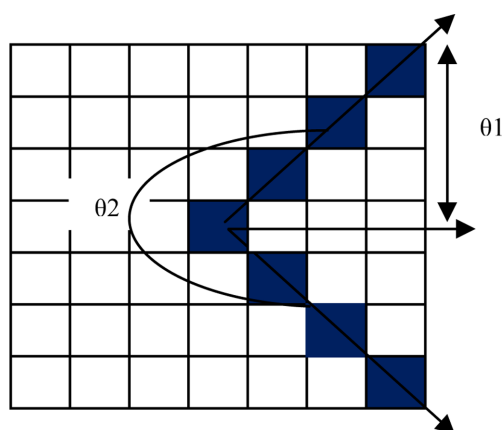


Fig. 5. Edge-Hinge feature calculation.

### B. Run Length

First, the grayscale image is converted into binary image by Otsu's thresholding. The normalized histogram of black pixels in the horizontal direction captured the intra-word and inter-word spacing in the handwritten document image. The information about the roundness in the handwriting is extracted and Run-Length was used as a feature. The example of black runs in horizontal direction for given image is as shown in Fig. 4. The size of feature vector is 100.

Width of black runs decides roundness in the handwriting image.

### C. Edge Hinge

The original image is a converted edge detected image using Sobel and followed by thresholding that generates a binary image in which only the edge pixels are activated. Considering each edge pixel in the middle of a square neighborhood as shown in Fig. 5, the presence of edge fragments (4 pixels wide) is checked. The histogram of

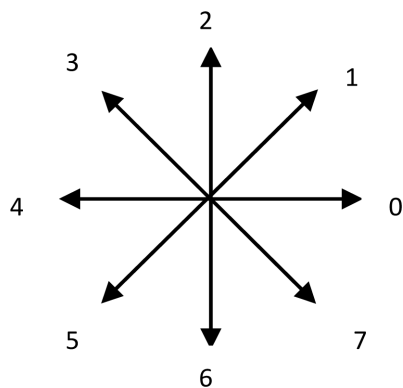


Fig. 6. Chain codes for 8 connected components.

all instances of two edge fragments (at  $\theta_1$  and  $\theta_2$ ) emerging from center pixel is counted. The PDF of  $p(\theta_1, \theta_2)$  capture slant and curvature in the handwriting is used as feature for writer identification and verification.

#### D. New Feature using Histogram of Chain Code

After capturing the directional information of a writer in the handwriting image, the writers are verified based on the pixel distribution information in 8 directions. Histograms are calculated using chain code of connected components in the handwritten document image.

First, the color image is converted to a grayscale and binarized. The contours are extracted from the negative of binarized image. The boundary detection algorithm is used to detect the contour. Then, each contour is represented using freeman chain code as shown in Fig. 6.

#### E. New Feature using Variance in Bins of Histogram of Chain Code

**1) Motivation:** The consistency in the writing style that varies from person to person has been observed and the natural variations or consistency in the writing is one of the factors that can be used to discriminate the writer. The natural variations in the writing are less for more consistent writing style and they are more for less consistent writing style as demonstrated in Fig. 7.

From Fig. 7, writers are found to have different degrees of writing variations and the style of writing a word (e.g., situation, problems) is found the same in image. The writing is more consistent in the first image (Fig. 7(a)) and in the second image (Fig. 7(b)) the style of writing a letter (i, l) / word (e.g., is, life) is found to be different. The writing is less consistent in the second image. Based on these observations, finding the variations in writing styles was determined to be significant and variance in bins of histogram of chain code was used as a feature for

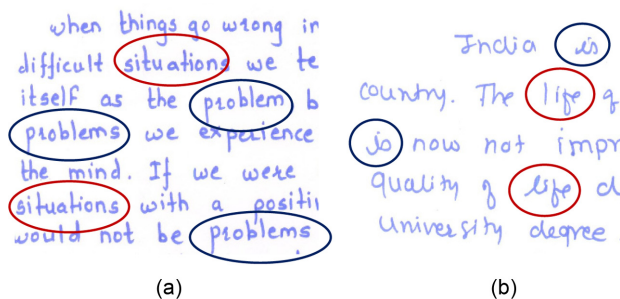


Fig. 7. Different writer samples having different writing styles: (a) handwriting image of writer 002 and (b) handwriting image of writer 101.

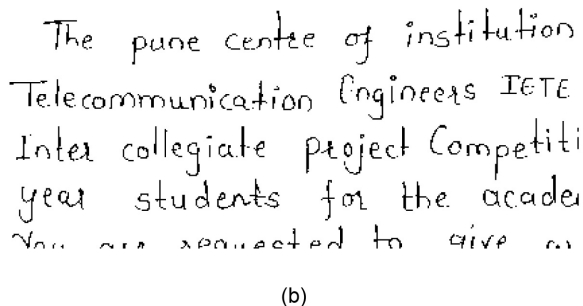
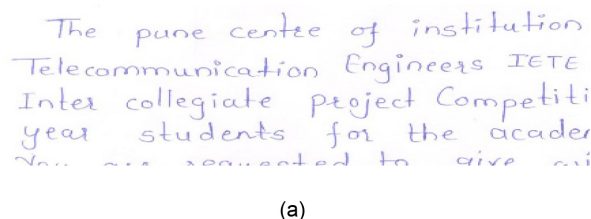


Fig. 8. Pre-processing step 1: thresholding of given image. (a) Input image and (b) binary image (size of original image is large. Hence, a small portion of original image is cropped, zoomed and shown in Fig. 8).

writer verification.

**2) Pre-processing:** First, the original color image is converted into grayscale and then converted into binary. Connected components (CCD) are extracted and then boundaries are detected for each CCD. Results of pre-processing are shown in Figs. 8 and 9. We used MATLAB 7.0.1 as a platform for coding.

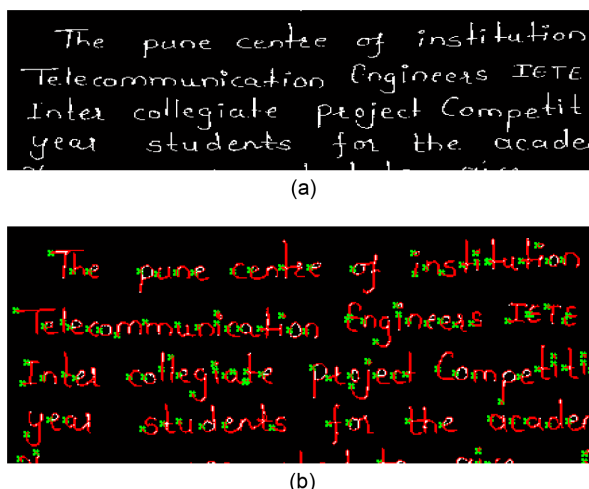
**3) Feature Extraction [18]:** After pre-processing, we calculated chain code as explained in Section III-D. Algorithm 1 explains chain code calculation.

---

#### Algorithm 1 Chain code calculation

---

- 1: compute dx, dy by a circular shift on cords arrays by 1 element
  - 2: check if boundary is close, if not cut last element
  - 3: check if boundary is 8-connected
  - 4: label boundary pixel using code shown in Fig. 6.
-



**Fig. 9.** Pre-processing step 2: boundary detection. (a) Negative of thresholded image shown in Fig. 8(b). (b) Boundary detection of connected components of negative image shown in Fig. 8(a). Boundaries are represented in red color and starting point of boundary is represented in green color (Size of original image is large. Hence a small portion of original image is cropped, zoomed and shown in Fig. 9).

To make the feature independent on starting point, we calculated histogram of chain code in 8 directions. The

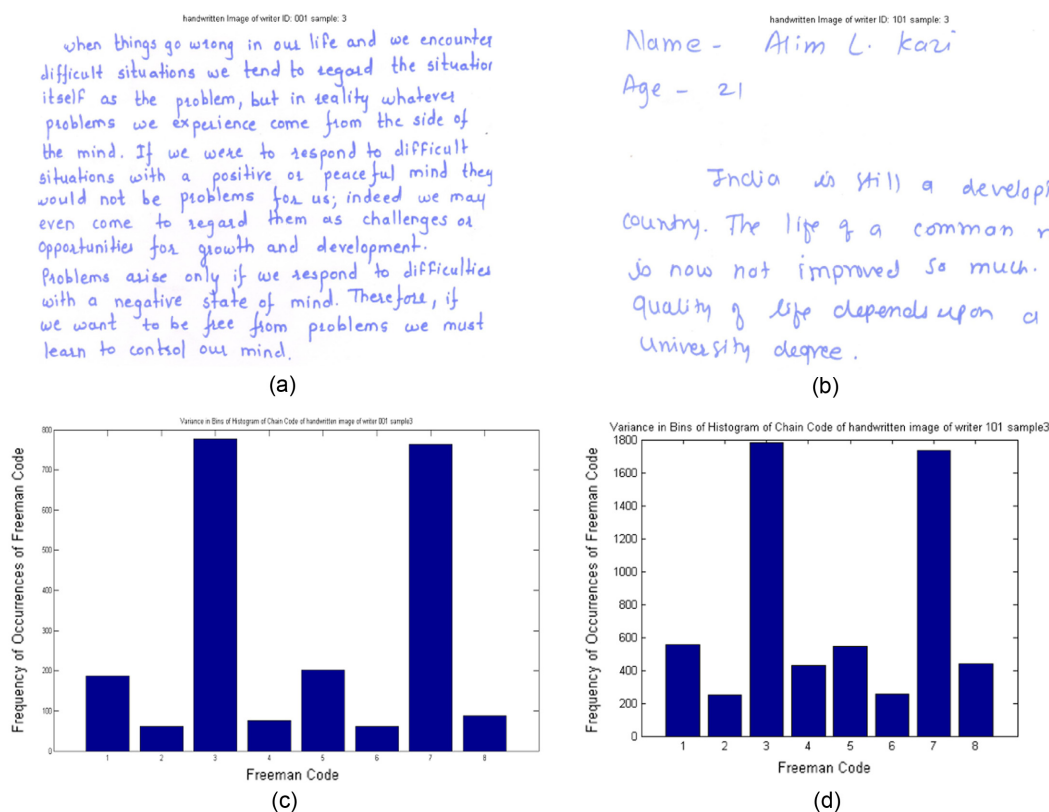
variance in bins of histogram is calculated and used as feature for writer verification. The variation in pixel distribution is calculated using variance parameter. The histogram of chain code of each CCD captured the information about the pixel distribution in 8 directions. To capture the information about the variation in each CCD in 8 directions, we calculated variance in bins of histogram of chain code as a feature and is shown in Fig. 10.

In Fig. 10, the first image has more consistency in writing compared to the second image. Therefore, the variance in bins of histogram of chain code is less in first image compared to second image, allowing variance feature to discriminate the handwriting samples effectively.

The feature is extracted from all the samples in the new dataset. After feature extraction, the system performed writer verification task.

#### IV. WRITER VERIFICATION

Writer verification is operated in two modes: training mode and testing mode. From the literature review, researchers (except Srihari in [19]) used IAM dataset, containing only two samples of each 650 writers to evaluate the system performance of their presented methods. Out of two samples, one sample of each writer was used



**Fig. 10.** Variance feature for different writer samples. (a) Handwriting image of writer 002, (b) handwriting image of writer 101, (c) variance in bins of histogram of chain code of image in (a), and (d) variance in bins of histogram of chain code of image in (b).

to train the system and one sample was used to test the system. We thought it inappropriate and therefore used two samples of each of the 1,000 writers to train and test the system. We considered one sample using ballpoint pen which gives minimum stroke width (1–2 pixel wide) and one sample using sketch pen which gives maximum stroke width (3–7) pixels.

### A. Training Mode

The training dataset includes two samples of each of the 1,000 writers including one sample using ballpoint pen and one sample using sketch pen. The city block L1 distance between the feature vector of ballpoint pen image and sketch pen image is used as a threshold  $T_w$  for writer  $w$ .

$$T_w = \sum_{i=1}^L |f_{wbi} - f_{wsi}| \tag{2}$$

$f_{wbi}$  is the  $i^{th}$  element of feature vector of handwriting sample using ballpoint pen image of writer  $w$ ,  $f_{wsi}$  is the  $i^{th}$  element of feature vector of handwriting sample using sketch pen image of writer  $w$ , and  $L$  is the length of feature vector. The threshold is calculated for all the writers in the dataset.

The calculated threshold value is stored along with feature vector for all the samples in the training dataset. Each template stored during the training mode consists of writer identification, feature vector, and threshold for writer. After the training mode, the system is entered into the testing mode.

### B. Testing Mode

In the testing mode, the claimed writer feature vector is compared with the feature vectors of remaining samples in the dataset. Comparison is accomplished on a one-to-one basis as shown in Fig. 11. The comparison results

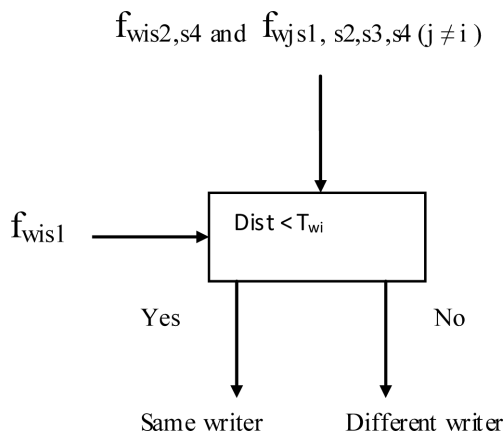


Fig. 11. Verification model.

into distances between same writer samples and distances between different writer samples. The distance is calculated using city block L1 distance measurement,

$$D_{u,k} = \sum_{i=1}^L |U_i - K_i| \tag{3}$$

where  $U$  is the feature vector of unknown sample (test sample 1),  $K$  is the feature vector of known sample (test sample 2) and  $L$  is the length of feature vector.

The new dataset includes four samples of 1,000 writers, and therefore has 2,000 distances of same writer samples and  $1,000 \times (999 \times 4) = 3,996,000$  distances of different writer samples.

If the distance ( $Dist$ ) between the two feature vectors is less than the threshold ( $T$ ) of claimed writer ( $w_i$ ) then the system answer is ‘Yes’ (same writer samples). If the distance ( $Dist$ ) between the two feature vectors is greater than the ( $T$ ) of claimed writer ( $w_i$ ) then the system answer is ‘No’ (different writer samples).

The system answer may be true or false. There are two types of errors. In case of similar handwritings of different person, distance between two feature vectors may be less than the threshold and system answer is Yes. This is the false acceptance error (FAR), where the false acceptance is calculated by presenting all the samples in the dataset. The performance is expressed in terms of %FAR.

$$\%FAR = \frac{\text{total number of falsely accepted samples}}{\text{maximum probability of false acceptance}} \times 100 \tag{4a}$$

The same writer samples may be quite different with different writing conditions. And, in such cases, the distance between the two feature vectors of same writers may be greater than the threshold and system answer is No. This is the false rejection error (FRR) and the false rejection is calculated by presenting all the samples in the dataset. The performance is expressed in terms of %FRR.

$$\%FRR = \frac{\text{total number falsely rejected samples} \times 100}{\text{maximum probability of false rejection}} \tag{4b}$$

Practically, with such few samples/writer, FAR cannot be equal to FRR. Therefore, to represent the performance in terms of single value, we calculated the average of both errors and called it average error rate (AER).

$$\%AER = \frac{FAR (\%) + FRR (\%)}{2} \tag{4c}$$

The performance is also expressed in terms of verification accuracy.

$$\%Verification\ Accuracy = 100 - AER (\%) \tag{4d}$$

The performance parameters (4a)–(4d) are calculated for all the samples in the dataset. The experimentation results are presented in the next section.



**Table 2.** Verification accuracy of most commonly used spatial domain methods of writer verification on new dataset under same and different stroke width conditions

Method	Verification accuracy (%)	
	Under same stroke widths (samples of ballpoint pen only/samples of sketch pen only)	Under different stroke width conditions (samples using ballpoint pen and sketch pen)
RL	60	39.74
E	80	59.53
EH	89	20

RL: run length (black runs in horizontal direction), E: entropy, EH: edge hinge.

## V. EXPERIMENTAL RESULTS AND DISCUSSIONS

The experimental results are evaluated using computer with i3 core processor, 2.40 GHz frequency, and 1.86 GB RAM.

### A. Performance Evaluation of the Most Commonly used Spatial Domain Methods

To study the effect of writing instrument on verification accuracy, we calculated accuracy using samples of ballpoint pen, samples of sketch pen and mixed samples in the dataset. The experimental result is presented in Table 2.

In Table 2, we compare two samples of same pen type (ballpoint pen vs. sketch pen), and the verification accuracy obtained is high as shown in column 2. But, when we compared one sample using ballpoint pen with another sample using sketch pen, we obtained very low verification accuracy as shown in column 3.

Entropy captured average ink distribution in handwritten document image, which varies at large scale for same writer samples with different ink width conditions. Therefore, the accuracy decreased to 60% with different ink width conditions.

Run-Length extracted information about inter-word and intra-word spacing in handwritten document image. The black run in horizontal direction depends on stroke widths in the document image, which depends on pen type used for writing. The Run-Length feature depends on the writing instrument and therefore we obtained low verification accuracy which failed under different ink width conditions.

The Edge-Hinge feature used edge detection using the Sobel operator as a pre-processing step for writer identification and verification. The handwriting slant was found to be +/- 15 degrees about the vertical axis. The Sobel operator failed to detect the pixel distribution in this range and resulted in a loss of useful information of pixel distribution and the accuracy is decreased. The Edge-Hinge feature failed to discriminate handwritings under different ink width conditions.

From the above experimental results, the Entropy, Run-

Length and Edge-Hinge features were concluded to have failed to discriminate the handwritings under different ink width conditions.

We tested histogram of chain code and variance in bins of histogram of chain code features on the new dataset.

### B. Performance Evaluation using Chain Code Features

The experimental result using histogram of chain code feature on our dataset is presented in Table 3.

In Table 3, the accuracy reported using histogram of chain code feature is 75.82% on the new dataset containing variations in ink distributions for same writer and different writer samples. The accuracy reported in a previous report [11] is 85% on IAM dataset, which we verified using our dataset considering samples of only ballpoint pen and sketch pen. The accuracy obtained was determined to be 84%.

From experimental results, the verification accuracy is decreased by 10% due to variations in ink widths in the handwritten image. Due to different stroke widths in two test samples, FAR is 11.67% but FRR is very high (36.70%) which is not acceptable. As a feature is extracted based on outer contour of CCD in handwriting image, there is large variation in same writer samples under different ink

**Table 3.** Performance of histogram of chain code feature on new dataset

Parameter	Document level	50% reduction in amount of text in document image
FAR (%)	11.67	13.42
FRR (%)	36.70	39.50
AER (%)	24.18	26.46
VA (%)	75.82	73.54
VT (s)	7.359000	7.338000
FET (s)	3847.578000	1791.609600

VA: verification accuracy, VT: verification time (second) for all samples in the new dataset, FET: feature extraction time for all samples in new dataset.

width conditions. Therefore, FRR is high, and thus the average error rate is high and average verification accuracy is low.

The effect of the amount of information on average verification accuracy was studied and the experimental result has shown that the reduction in amount of text by 50%, increased %FAR by 1.75, %FFR by 2.80, %AER by 2.28, and decreased %verification accuracy by 2.28. The greater the amount of information contained in the document image, the higher the accuracy.

The experimental result using variance in bins of histogram of chain code feature on the new dataset is presented in Table 4.

The style of a writer is greatly affected by the type of pen used for writing. Cursive is easy to write using ball-point pen but difficult using a sketch pen. The degrees of writing variations are greater for the same writer samples with different ink width conditions. Therefore, the FRR is very high, which is not acceptable, and thus the average error rate is high and average verification accuracy is low.

In Table 4, the reduction in amount of text by 50% increased FAR by only 0.22%, FFR by 2.40% and decreased accuracy only by 1.31% on the new dataset containing variations in ink distributions for same writer and different writer samples.

The performance comparison of histogram of chain

**Table 4.** Performance of variance in bins of histogram of chain code feature on new dataset

Parameter	Document level	50% reduction in amount of text in document image
FAR (%)	16.63	16.85
FRR (%)	38.75	41.15
AER (%)	27.69	29.00
VA (%)	72.31	71.00
VT (s)	7.891000	7.125000

VA: verification accuracy, VT: verification time (second) for all samples in the new dataset.

**Table 5.** Performance comparison of histogram of chain code feature and variance in bins of histogram of chain code feature (document level)

Parameter	Histogram of chain code feature	Variance in bins of histogram of chain code feature
FAR (%)	11.67	16.63
FRR (%)	36.70	38.75
AER (%)	24.18	27.69
VA (%)	75.82	72.31
VT (s)	7.359000	7.891000

VA: verification accuracy, VT: verification time (second) for all samples in the new dataset.

code and variance in bins of histogram of chain code is presented in Table 5.

Referring to Table 5, the histogram of chain code feature performed well compared to the variance in bins of histogram of chain code feature.

Accuracy obtained using histogram of chain code feature (75.82%) and variance in bins of histogram of chain code feature (72.31%) is high compared to Entropy (60%), Run-Length (40%) and Edge-Hinge features (20%).

To make the feature independent of ink width variations, the chain code based features are extracted from outer contour of connected components in the handwritten image. Therefore, this method appeared to have higher performance than the other methods with different ink width conditions.

We also studied the effect of dataset size on verification accuracy presented in Table 6.

Referring to Table 6, as we increased the size of the dataset from 200 writers to 800 writers, the verification accuracy decreased from 78.09% to 75.82% using histogram of chain code feature and 76.69% to 72.31% using variance in bins of histogram of chain code feature. The problem with the new spatial domain feature is that the FRR is very high which is not acceptable.

The performance comparison of all presented methods is given in Table 7.

Referring to Table 7, the histogram of chain code out-

**Table 6.** Effect of dataset size on verification accuracy (new dataset)

No. of writers	Verification accuracy (%)	
	Histogram of chain code feature	Variance in bins of histogram of chain code feature
200	78.09	76.69
400	76.76	74.78
600	76.01	72.23
800	76.66	73.02
1,000	75.82	72.31

**Table 7.** Performance comparison of presented spatial domain methods

Method	Accuracy (%)	Testing time (s)	Feature length (elements)
HCC	75.82	7.359000	8
VBHCC	72.31	7.891000	8
E	59.53	0.625000	1
RL	39.74	156.000	100
EH	20	358.870000	12

HCC: histogram of chain code, VBHCC: variance in bins of histogram of chain code, E: entropy, RL: run length (black runs in horizontal direction), EH: edge hinge

performed the other methods with different ink width conditions. The advantage of new feature is the use of less memory space, feature extraction time and verification time because of small size of feature vector (only 8 elements).

## VI. CONCLUSIONS AND FEATURE SCOPE

In this paper, we address the emerging issue of writer verification using different ink width conditions. We evaluated the performance of existing spatial domain methods of writer identification and verification under different ink width conditions. The experimental results show that most commonly used features such as Entropy, Run-Length, and Edge-Hinge are not suitable under different ink width conditions.

To make the system independent of ink width, we extracted contour based features using a chain code method. To improve accuracy, we considered histograms of chain code and variance in bins of histograms of chain code as new features using different ink width conditions. The writers are verified based on pixel distribution and variation in pixel distribution in handwritten English document images. The system was tested on our newly created dataset of 1,000 writers contributing 4 samples from each writer.

Experimental results have shown that the histogram of chain code outperformed other methods such as Entropy, Run-Length and Edge-Hinge, with a verification accuracy of 75.82% on the new dataset. The verification time is 7.359000 seconds for all samples in the dataset using MATLAB 7.0.1 on an i3 core processor with 2.64 GHz clock and 1.86 GB RAM. The feature vector length is only 8 elements. The FRR was observed to be very high at 36.70%—which is not acceptable.

Based on the experimental results, we conclude that the spatial domain features for writer identification and verification provided low FAR but FRR is high under different ink width conditions.

To improve accuracy further, we propose multiple feature combinations using spatial domain techniques for writer identification and verification. The transform domain based features for writer identification and verification were also studied using different ink width conditions. The performance of spatial domain and transform domain techniques can also be compared to select the best feature(s) for writer identification and verification with different ink width conditions.

## REFERENCES

1. U. V. Marti and H. Bunke, "The IAM-database: an English sentence database for offline handwriting identification and verification," *International Journal of Document Analysis*

- and Identification and verification*, vol. 5, no. 1, pp. 39-46, 2002.
2. S. L. Kore and S. D. Apte, "The current state of art-the handwriting a behavioral biometric for writer identification and verification," in *Proceedings of the ACM International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Chennai, India, 2012, pp. 925-930.
3. B. Arazi, "Handwriting identification by means of run-length measurements," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 7, no. 12, pp. 878-881, 1977.
4. M. Bulacu and L. Schomaker, "Text-independent writer identification and verification using textural and allographic features," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 701-717, 2007.
5. L. Schomaker and M. Bulacu, "Automatic writer identification using connected-component contours and edge-based features of upper-case western script," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 6, pp. 787-798, 2004.
6. M. L. Bulacu, "Statistical pattern identification and verification for automatic writer identification and verification," Ph.D. dissertation, State University of Buffalo, New York, NY, 2007.
7. A. Bensefia, T. Paquet, and L. Heutte, "A writer identification and verification system," *Pattern Recognition Letters*, vol. 26, no. 13, pp. 2080-2092, 2005.
8. N. Akhter, "Script independent offline writer identification using handwriting style," in *Proceedings of 3rd International Conference on Crime Detection and Prevention (ICDP'09)*, London, 2009, pp. 1-5.
9. A. A. Brink, J. Smit, M. L. Bulacu, and L. Schomaker, "Writer identification using directional ink-trace width measurements," *Pattern Recognition*, vol. 45, no. 1, pp. 162-171, 2012.
10. I. Siddiqi and N. Vincent, "Text independent writer recognition using redundant writing patterns with contour-based orientation and curvature features," *Pattern Recognition*, vol. 43, no. 11, pp. 3853-3865, 2010.
11. I. Siddiqi and N. Vincent, "A set of chain code based features for writer recognition," in *Proceedings of 10th International Conference on Document Analysis and Recognition*, Barcelona, Spain, 2010, pp. 981-985.
12. Z. Huang, D. Wang, and Y. Lu, "Writer identification using differential chain code and grid features," in *Foundations of Intelligent Systems*, Heidelberg: Springer, 2014, pp. 647-656.
13. H. Ding, H. Wu, X. Zhang, and J. Chen, "Writer identification based on local contour distribution feature," *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 7, no. 1, pp. 169-180, 2014.
14. S. S. Ram and M. E. Moghaddam, "A Persian writer identification method using swarm-based feature selection approach," *International Journal of Biometrics*, vol. 6, no. 1, pp. 53-74, 2014.
15. A. Chaabouni, H. Boubaker, M. Kherallah, H. El-Abed, and A. M. Alimi, "Static and dynamic features for writer identification based on multi-fractals," *International Arab Journal of Information Technology*, vol. 11, no. 4, pp. 416-424, 2014.
16. M. N. Abdi and M. Khemakhem, "A model-based approach

to offline text-independent Arabic writer identification and verification,” *Pattern Recognition*, vol. 48, no. 5, pp. 1890-1903, 2015.

17. I. Siddiqi and N. Vincent, “Stroke width independent feature for writer identification and handwriting classification,” in *Proceedings of the 11th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, Montreal, Canada, 2008, pp. 1-6.
18. S. L. Kore and S. D. Apte, “Ink width independent global features for writer verification,” in *Proceedings of 2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Mysore, India, 2013, pp. 1770-1775.
20. H. Arora, S. Lee, S. N. Srihari, and S. H. Cha, “Individuality of handwriting,” *Journal of Forensic Science*, vol. 47, no. 4, pp. 1-17, 2002.
19. A. Shivram, C. Ramaiah, S. Setlur, and V. Govindaraju, “IBM\_UB\_1: a dual mode unconstrained English handwriting dataset,” in *Proceedings of 12th International Conference on Document Analysis and Recognition (ICDAR)*, Washington, DC, 2013, pp. 13-17.



### Sharada Laxman Kore

Sharada Laxman Kore received Diploma in Industrial Electronics Engineering from Walchand College of Engineering, Sangli, under Board of Technical Education, Maharashtra, India in the year 1995. She received B.E. and M.E. degrees in Electronics Engineering from Walchand College of Engineering, Sangli, under Shivaji University, Kolhapur, Maharashtra, India in 1998 and 2000, respectively. Presently, she is pursuing her Ph.D. degree from Bharati Vidyapeeth University College of Engineering, Pune, Maharashtra, India under the guidance of Prof. Dr. Shaila Apte. She has 15 years of experience in teaching. Currently, she is holding the position of Associate Professor in E&TC Dept. at Bharati Vidyapeeth's College of Engineering for Women, Pune, Maharashtra, India. She is a member of executive body of IETE local centre, Pune, India for the period 2014-2016. She is a life member of Institution of Electronics and Telecommunication Engineers and Indian Society for Technical Education. Her area of interest includes Image Processing, Pattern Recognition and Writer Identification and Verification.



### Shaila Dinkar Apte

Shaila Dinkar Apte received M.Sc. degree in Electronics from Mumbai University, Maharashtra, India in 1976 with the first rank. She received M.E. and Ph.D. degrees in Electronics Engineering from Walchand College of Engineering, Sangli under Shivaji University, Kolhapur, Maharashtra, India in 1991 and 2001, respectively. She is currently working as a Professor at Rajarshi Shahu College of Engineering, Pune, Maharashtra, India. She has formerly been an Assistant Professor in Walchand College of Engineering, Sangli, Maharashtra, India for 27 years; a member of the Board of Studies at Shivaji University and a Principal Investigator for a research project sponsored by the Armament Research and Development Establishment (ARDE), New Delhi, Maharashtra, India. She has teaching experience of 35 years in Electronics Engineering. Under her guidance, more than 60 candidates have completed their M.E. dissertations and seven candidates have completed their Ph.D. theses and five are pursuing. She has a patent published to her credit related to the generation of the mother wavelet from a speech signal, 77 paper publications, and 3 books with Wiley India Publications title DSP, ADSP, Speech and Audio processing. She has been teaching Digital Signal Processing for the last 25 years. She is a regular reviewer of Speech Processing journal of Springer and Signal processing journal of Elsevier.