Recognition of Isolated Handwritten Characters in Gurmukhi Script

Dharamveer Sharma
Lecturer,
Department of Computer Science, Punjabi University, Patiala

Puneet Jhajj
MTech (Computer Science),
Department of Computer Science, Punjabi University, Patiala

ABSTRACT
Isolated handwritten character recognition has been the subject of intensive research during last decades because it is useful in wide range of real world problems. It also provides a solution for processing large volumes of data automatically. Work has been done in recognizing handwritten characters in many languages like Chinese, Arabic, Devnagari, Urdu and English. The work presented in this thesis, focuses on the problem of recognition of isolated handwritten characters in Gurmukhi script. The whole process consists of two stages. The first, feature extraction stage analyzes the set of isolated characters and selects a set of features that can be used to uniquely identify characters. The performance of recognition system depends heavily on what features are being used.
The selection of stable and representative set of features is the heart of recognition system. The feature extraction method Zoning, is used for extracting features of the character under consideration in this problem. In Zoning method, the frame containing the character is divided into several overlapping or non-overlapping zones and the densities of object pixels in each zone are calculated. Densities are used to form a representation. The final, classification stage is the main decision making stage of the recognition system. It uses features extracted in the feature extraction stage to identify the character. K-Nearest Neighbor and Support Vector Machine are the two classifiers used for identifying the character in the problem. In k-nearest classification method, the Euclidean distance between the test point and all the reference points is calculated in order to find K nearest neighbors, and then the obtained distances are ranked in ascending order and the reference points corresponding to the k smallest Euclidean distances are taken. The Support Vector Machine (SVM) is learning machine with very good generalization ability. SVM implements the Structural Risk Minimization Principal which seeks to minimize an upper bound of the generalization error. An SVM classifier discriminates two classes of feature vectors by generating hyper-surfaces in the feature space, which are "optimal" in a specific sense that is the hyper-surface obtained by the SVM optimization is guaranteed to have the maximum distance to the "nearest" support vectors. SVM operate on kernel evaluations of the feature vectors. An annotated sample image database of isolated handwritten characters in Gurmukhi script has been prepared which has been used for training and testing of the system.

1. INTRODUCTION
Optical character recognition, abbreviated as OCR, is the process of converting the images of handwritten, typewritten or printed text (usually captured by a scanner) into machine-editable text or computer processable format, such as ASCII code. Computer systems armed with OCR system improve the speed of input operations, reduce data entry errors, reduce storage space required by paper documents and thus enable compact storage, fast retrieval, scanning corrections and other file manipulations. OCR have applications in postal code recognition, automatic data entry into large administrative systems, banking, automatic cartography, 3D object recognition, digital libraries, invoice and receipt processing, reading devices for blind and personal digital assistants. OCR includes essential problems of pattern recognition. Accuracy, flexibility and speed are the three main features that characterize a good OCR system. OCR aims at enabling computers to recognize optical symbols without human intervention. This is accomplished by searching a match between the features extracted from a given symbol's image and the library of image models. The basic process of OCR Systems is shown in Figure1.

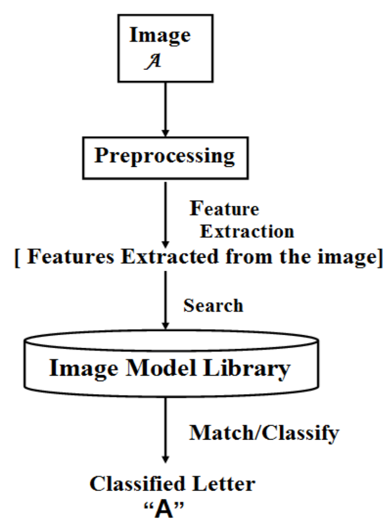


Figure 1: The basic process of an OCR System

The process of optical character recognition of any script can be broadly broken down into 6 stages as shown in Figure2:

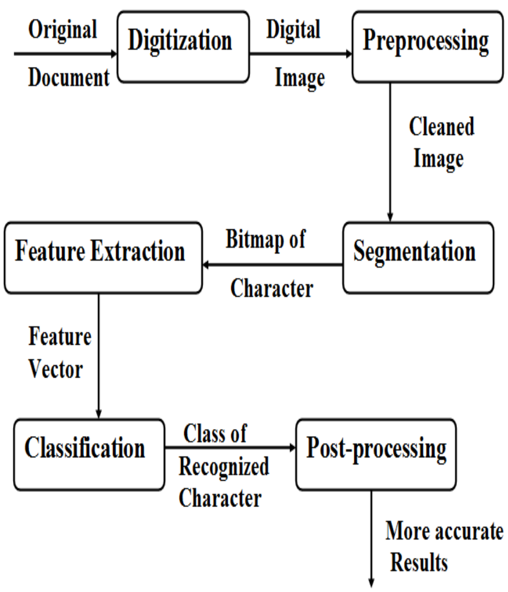


Figure 2: Block diagram of OCR system

1. Digitization: Digitization produces the digital image, which is fed to the pre-processing phase.
2. Preprocessing: The image produced by digitization may carry some unwanted noise. The preprocessing stage takes in a raw image, reduces noise and distortion, removes skewness and performs skeltonizing of the image. After preprocessing phase, we have a cleaned image which goes to the segmentation phase.
3. Segmentation: The segmentation stage takes in the image and separates the different logical parts, like lines of a paragraph, words of line and characters of a word.
4. Feature Extraction: After segmentation, set of features is required for each character. In feature extraction stage every character is assigned a feature vector to identify it. This vector is used to distinguish the character from other characters. Various feature extraction methods are designed like zoning, PCA, Central moments, structural features, Gabor filters and Directional Distance Distribution. Feature extraction is the process of selection of the type and the set of features
5. Classification: Classification is the main decision making stage of OCR system. It uses the features extracted in the previous stage to identify the text segment according to preset rules. Many type of classifiers are applicable to OCR like K-nearest neighbour, Neocognitrons and SVM.
6. Post processing: The output of classification may contain some recognition errors. Post-processing methods remove these errors by making use of mostly two methods namely, dictionary lookup and statistical approach.

1.1 Types of Handwriting recognition

Handwriting recognition is broken into two different types depending on how the handwritten data is presented to the recognition system.

1. Online Handwriting Recognition:
   In online recognition systems, the computer recognizes the symbols as they are drawn. Online recognition basically goes along the writing process.

2. Offline Handwriting Recognition:
   Offline handwriting recognition is performed after writing is complete. Offline handwriting recognition is performed after the writing or printing is complete.

1.2 Introduction to Gurmukhi script

Gurmukhi script is used primarily for Punjabi language, which is the world's 14th most widely spoken language. Some of the properties of Gurmukhi script are:

- Gurmukhi script is cursive and the character set consist of 41 consonants,9 vowels, 3 sound modifiers(semi-vowels) and 3 half characters, which lie at the feet of consonants. The Character get of Gurmukhi script is described in Figure3.

Vowel Carriers:					
ੳ	ਅ	ੲ			
Consonants:					
ਸ	ਹ				
ਕ	ਖ	ਗ	ਘ	ਙ	
ਚ	ਛ	ਜ	ਝ	ਞ	
ਟ	ਠ	ਡ	ਢ	ਣ	
ਤ	ਥ	ਦ	ਧ	ਨ	
ਪ	ਫ	ਬ	ਭ	ਮ	
ਯ	ਰ	ਲ	ਵ	ੜ	
ਸ਼	ਖ਼	ਗ਼	ਜ਼	ਫ਼	ਲ਼

Vowels:								
ਾ	ਿ	ੀ	ੁ	ੂ	ੇ	ੈ	ੋ	ੌ
Semi-vowels:								
ੰ	ੱ	ਂ						
Half Characters:								
੍ਹ	੍ਰ	੍ਵ						

Figure 3: Character set of Gurmukhi Script

- Most of the Gurmukhi characters have a horizontal line at the upper part. The characters of words are connected mostly by this line called head line and so there is no vertical inter-character gap in the letters of a word. For example:

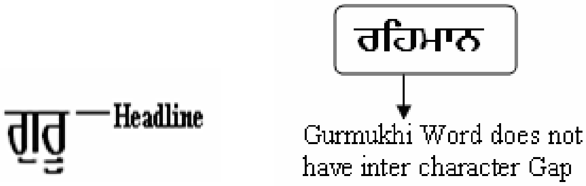


Figure 4: Headline and missing inter-character gap

- A word in Gurmukhi script can be partitioned into three horizontal zones, as shown in Figure5. The upper zone denotes the region above the head line, where vowels reside, while the middle zone represents the area below the head line where the consonants and some sub-parts of vowels are present. The middle zone is the busiest zone. The lower zone represents the area below middle zone where some vowels and certain half characters lie in the foot of consonants. But there is no concept of upper and lower zones in Gurmukhi digits.

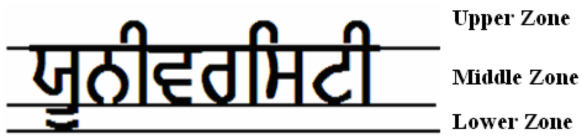


Figure 5: Horizontal Zones

- The bounding boxes of 2 or more characters in a word may intersect or overlap vertically. For example bounding boxes of ਪ and ਂ intersect

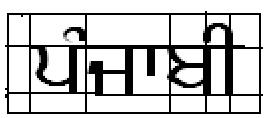


Figure 6: Intersecting/overlapping characters

- There are lots of topologically similar character pairs in Gurmukhi script. Some similar pairs are
  ਖ and ਖ਼, ਵ and ੜ, ਖ and ਖ਼, ਤandਡ etc

1.3 Objectives of Research:

The objectives of the proposed study are outlined as follows:

1. To develop algorithms and procedures which extract a set of features from the images of isolated handwritten characters in Gurmukhi script and use the set of features to recognize the characters.
2. To develop software module, based on the above algorithms, which will recognize the isolated handwritten characters in Gurmukhi script.

1.4 Assumptions

We have considered following assumptions while developing the method for recognition of isolated handwritten characters in Gurmukhi Script:

- We have considered pre-segmented isolated handwritten alphabets of Gurmukhi script.
- Character images are noise free.
- Characters are considered without stripping their headline, if they have any, to preserve the information about presence or absence of headline.
- We have considered 41 consonants of Gurmukhi alphabet only.

1.5 Problem Description

Recognition of isolated handwritten characters is the process of identifying individual characters. It is useful in wide range of real world problems like documentation analysis, mailing address interpretation, bank check processing, signature verification, documentation verification and many others. Due to applications of recognition, it is one of the most challenging areas of pattern recognition. It has been topic of research for a
long period of time. Work has been done in recognizing handwritten Chinese, Arabic, Devnagari, Urdu and English characters, recognizing handwritten numerals and handwritten digits. Here the problem is to recognize the isolated handwritten characters in Gurmukhi script.

The major difficulties are:

1. The variability of writing styles, both between different writers and between separate examples from the same writer overtime. For example:

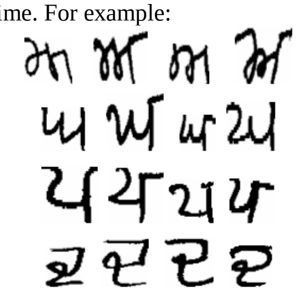


Figure 7: Varying writing styles

2. The similarity of some characters. Table1 shows examples of similar characters

Table 1: Similar Characters

ਖ and ਖ਼	ਵ and ੜ	ਜ andਜ਼
ਗ and ਘ	ਤ and ਝ	ਬ and ਥ
ਨ andਠ	ਤandਡ	ਥ and ਖ
ਫ and ਫ਼	ਟand ਦ	ਪ and ਧ
ਅ andਘ	ਹ and ਰ	ਪ and ਖ
ਸ andਮ	ੲ and ਟ	ਗ andਗ਼
ਸ and ਸ਼	ਵ andੜ	ਚ andਦ
ਲ and ਲ਼		

3. The possible low quality of the text image. For example:

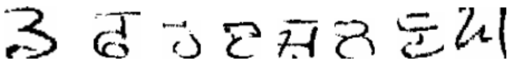


Figure 8: Low quality of text images

4. The unavoidable presence of background noise and various kinds of distortions (such as poorly written, degraded, or overlapping characters) can make the recognition process even more difficult. For example:

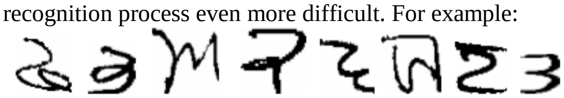


Figure 9: Poorly written characters

In the problem of recognition of isolated handwritten characters the input is isolated characters. Word segmentation provides isolated characters. Characters can be in upper zone, middle zone or lower zone.
Some of the examples are shown in figures below:
Upper zone:

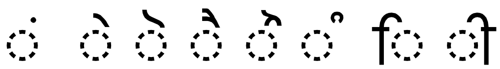


Figure 10: Upper zone characters

But in case of vowels ੀ and ਿ, lie in middle zone and lie in upper zone.

Middle zone:

Figure 11: Middle zone characters

But in case of character ੳ, lies in upper zone and lies in middle zone. Vowel ਾ lies in middle zone

Lower zone:

Figure 12: Lower zone characters

The organization of this paper is: Section 2 includes literature survey, section 3 includes proposed method, and section 4 includes results and discussion.

2. LITERATURE SURVEY

A lot of research has been done on the recognition of handwritten characters in the recent years, resulting in a number of proposed pattern recognition techniques. Lazzerini and Marcelloni [1], presents EYE, a fuzzy logic based classifier for recognition of isolated handwritten characters. EYE is based on a new linguistic classification method. The method describes characters in terms of linguistic expressions and adopts a purposely defined operator to compare these expressions. Hanmandlu et al. [2], presents an innovative approach called box method for feature extraction for the recognition of handwritten characters. In this approach, the character image is partitioned into a fixed number of sub images called boxes. The features consist of normalized vector distance and angle from each box to a fixed point. The recognition schemes used are back propagation neural network (BPNN) and fuzzy logic. The recognition rate is found to be around 100% with the fuzzy based approach on the standard database. Zhang et al. [3], proposed a handwritten character recognition feature based on the combination of gradient feature and coefficients of wavelet transform. The gradient feature represents local characteristic of a character image properly, but it is sensitive to the deformation of handwritten character. The wavelet transform represents the character image in multiresolution analysis and keeps adequate global characteristic of a character image in different scales. Gary and Joe [4], proposed a new method to measure the similarity between two fuzzy attributed graphs of a known character class and an unknown character. In the recognition stage, when this similarity measure is applied, an input character can be correctly classified. Liolios et al. [5], presents a system capable of recognizing isolated handwritten characters using shape transform method. The shape transform approach is based on the calculations the cost of transforming the image of a given character into that of another, thus taking into account local geometrical similarities and differences. Tou and Gonalez [6], designed a handwritten character recognition system making use of topological feature extraction and multilevel decision making. By properly specifying a set of easily detectable topological features, the system classifies the handwritten characters into primary categories. The recognition system consists of two stages, with the final classification accomplished by a secondary stage that performs local analysis on the characters in each primary category. The recognition system consists of two stages: global recognition, followed by local recognition. Heutte et al. [7], present a new feature vector for the recognition of handwritten characters which combines the strengths of both statistical and structural feature extractors and provides a wide range of identification clues. In this the author investigates the application of structural features to the statistical recognition of handwritten characters. In order to feed structural features to statistical classifier, parameterization of the structural features has been proposed and allows to get rid of the most common used transformation of these features into binary values. . Araki et al. [8], have proposed a new character recognition algorithm using a Bayesian filter. Handwritten character image that is read by a scanner is used. The proposed method consists of three steps, preprocessing, learning and recognition step. In the preprocessing step, threshold processing for the scanner image is carried out and the image size is normalized. Then, to obtain the appearance probability of a black pixel in each pixel of a certain character image, the appearance count of a black pixel of learning data is determined in the learning step. Finally, the character recognition is executed to calculate the probability that the given character image is the same as a learning character by Bayesian filter using the appearance probability. Experimental results have clarified that the proposed method has over a 90% recognition rate even though it uses only a few learning data. Hu and Yan [9], presented a structural method for describing both printed and handwritten characters. The character is decomposed into primitives by detecting feature points, and each primitive is described by a primitive code. The topological information of a character is represented by a global code. The global code and the primitive codes describe each character clearly and effectively. U. Pal et al.[10] has presented a system for recognition of handwritten characters of Devnagri Script. The features used for recognition are mainly based on directional information obtained from the arc tangent of the gradient. To get the feature, at first a 2x2 man filtering is applied 4 times on the gray level image and a non-linear normalization is done on the image. The normalized image is segmented to 49x49 blocks and a Roberts filter is applied to obtain gradient image. The arc tangent of gradient is initially quantized into 32 directions and the strength of the gradient is accumulated with each of the quantized direction. Finally, the blocks and the directions are down sampled using Gaussian filter to get 392 dimensional feature vectors. A modified quadratic classifier is applied on these features for recognition.

3. PROPOSED METHOD

In this paper a system to recognize isolated handwritten characters in Gurmukhi characters has been developed. In this images of isolated handwritten characters are provided as input. Then feature extraction method extracts the features of the characters and finally, classifiers identify the characters using the features extracted by feature extractors. In this the feature extraction method, Zoning and two classification methods, k-nearest neighbor, SVM (support vector machines) have been used and compared.

3.1 Feature extraction method:

The performance of a character recognition system depends heavily on what features are being used. Selection of a feature extraction method is probably the single most important factor in achieving high recognition. The extracted features should be able to identify each character set uniquely. There should be large variations in the features of different character sets. For several decades many kinds of features have been established and their test performances on standard database have been reported. Different types of features can be extracted depending on the representation forms of characters

3.1.1. Zoning

The frame containing the character is divided into several overlapping or non-overlapping zones and the densities of object pixels in each zone are calculated. Density is calculated by finding the number of object pixels in each zone and dividing it by total number of pixels. Densities are used to form a representation. For binary images, value of each pixel is either 1 or 0. We have considered pixels having value BLACK (0) as object pixels. This feature is extracted from the scaled (normalized) character matrix of the character.

The original character image (matrix) is first scaled to Normalized window of size 48*48.

Figure 13: Original Image of Gurmukhi character sassa (ਸ)

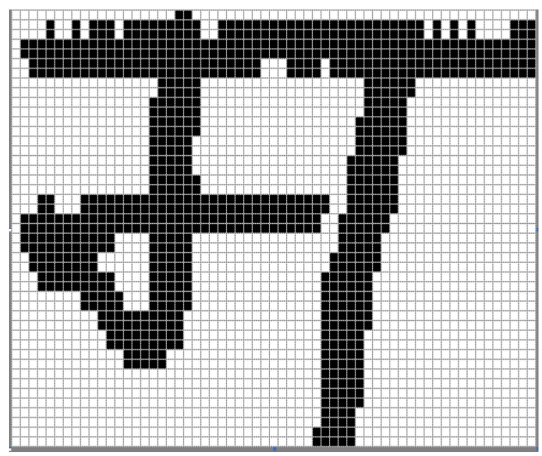


Figure 14: Scaled Image of Gurmukhi Character sassa (ਸ)

The Zoning feature set consists of 64 values. The values in feature vector are normalized in the range 0 to 1. Normalization is done by dividing all the values by the largest value in the feature set.

3.2 Classification methods:

Classification stage uses the features extracted in the feature extraction stage to identify the text segment. It is concerned with making decisions concerning the class membership of pattern in question. Classification determines the region of feature space in which an unknown pattern falls.
The task here is to design model using training data which can classify the unknown pattern based on that model. For training purposes we have used isolated Gurmukhi characters written in different forms. Feature vector for all training data is produced and stored in files

3.2.1 K-Nearest Neighbour:

The k-nearest neighbor (k-nn) approach attempts to compute a classification function by examining the labeled training points as nodes or anchor points in the n-dimensional space, where n is feature size. We calculate the Euclidean distance between the test point and all the reference points in order to find K nearest neighbors, and then rank the obtained distances in ascending order and take the reference points corresponding to the k smallest Euclidean distances. A test sample is then attributed the same class label as the label of the majority of its K nearest (reference) neighbors. Euclidean distance is the straight line distance between two points in n-dimensional space. The Euclidean distance between an input feature vector X and a library feature vector C is given by following equation:

$$\sqrt{\sum_{i=1}^{N}(C_i - X_i)(C_i - X_i)}$$

Where C_i is the ith library feature and X_i is the ith input feature and N is the number of features used for classification. The class of the library feature vector producing smallest Euclidean distance, when compared with the library input feature vector, is assigned to the input character. The k-nn is more general than nearest neighbor. In other way, nearest-neighbor is special case of k-nn, where k=1.

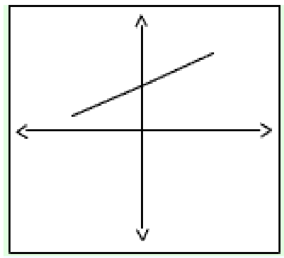
3.2.2. SVM (Support Vector Machines):

SVM (Support Vector Machine) is a useful technique for data classification [11, 12, 13]. The Support Vector Machine (SVM) is learning machine with very good generalization ability, which has been applied widely in pattern recognition, regression estimation, isolated handwritten character recognition, object recognition speaker identification, face detection in images and text categorization. SVM implements the Structural Risk Minimization Principal which seeks to minimize an upper bound of the generalization error. SVM is a kind of learning machine whose fundamental is statistics learning theory. An SVM classifier discriminates two classes of feature vectors by generating hyper-surfaces in the feature space, which are "optimal" in a specific sense that is the hyper-surface obtained by the SVM optimization is guaranteed to have the maximum distance to the "nearest" training examples, the support vectors. On the binary linear separable case, SVM determines the optimal hyper-plane through maximizing the margin between the separating hyper-plane and the data, subjecting to the constraint of classifying the training samples correctly. This can be regarded as an approximate implementation of the structure risk minimization (SRM) principle. SVM can be extended easily to the nonlinear classifier by projecting the data into a high dimensional feature space. It only needs to compute the dot product in the input space rather than in the feature space via constructing a certain kernel function, even need not know the mapped patterns explicitly. SVM operate on kernel evaluations $K(P_i,P_j)$ of the feature vectors P_i or P_j. Variant learning machines are constructed according to the different kernel functions and thus construct different hyperplanes in the feature space.

Different types of kernel functions of SVM:

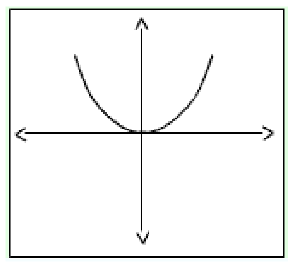
Linear kernel:

$$K(x_i, x_j) = x_i^T x_j$$



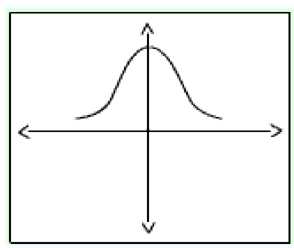
Polynomial Kernel:

$$K(x_i, x_j) = (x_i^T x_j + 1)^d$$



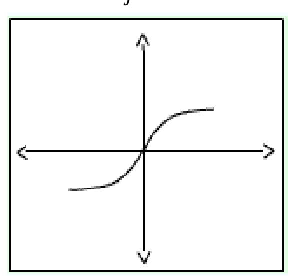
RBF kernel:

$$K(x_i, x_j) = \exp(-\|x_i - x_j\|^2 / 2\sigma^2)$$



Sigmoid Kernel:

$$K(x_i, x_j) = \tanh(\gamma x_i^T x_j - \Theta)$$



4. RESULTS AND DISCUSSIONS:

An annotated sample image database of isolated handwritten characters in Gurmukhi script has been prepared. The database contains name of source image, size of image and character value of the image. For storage of image data XML format has been used. We have experimented the system on 2050 images of Gurmukhi characters contained in the database. The system is analyzed using different combinations of feature extraction methods and classification methods. We have used 3075 images to train the system and 2050 for testing. The recognition accuracy obtained by using different combinations of feature extraction methods and classifiers is given in the Table 2. The recognition accuracy is obtained by dividing the correctly recognized characters to total number of character images which are actually present in the database.

Table 2: Performance of different combinations of feature extraction method and classification techniques

Feature extraction	Classifier	Total Images	Correctly recognized	Recognition Accuracy
Zoning	KNN	2050	1489	72.54%
Zoning	SVM (Linear Kernel)	2050	1490	72.68%
Zoning	SVM (Poly. Kernel)	2050	1497	73.02%
Zoning	SVM (RBF Kernel)	2050	1493	72.83%

Zoning with SVM (Polynomial kernel) gives the best results of all the combinations of feature extraction methods and classification methods as is evident from the Table2.

4.1 Reasons of Failure:

Sometimes the characters are also wrongly classified. It happens due to many reasons like

- Character is not properly written.
- Some characters have similar topological structures.
- Sometimes the headline is not properly written.
- Low quality images(images are blurred, headline is broken etc)

System sometimes confuses the character with some other character and does not recognize it correctly.

In the Table3 the confusion matrix caused when Zoning and SVM (Polynomial Kernel) have been used is given. We have taken 2050 images for 50 images of each of the 41 characters.

The characters ੳ, ਗ, ਘ, ਝ, ਨ, ਲ are recognized with higher accuracy and characters ਅ, ਢ, ਧ, ਫ are recognized with least accuracy as is evident from the Table3.

Table 3: Confusion matrix for Gurmukhi Alphabet

Character	Confused with characters									
ੳ	ੳ 93.81%	ਚ 1.70%	ਡ 0.80%	ਵ 2.89%	ੜ 0.80%					
ਅ	ਅ 61.37%	ਘ 38.63%								
ੲ	ੲ 75.33%	ਝ 3.41%	ਟ 5.44%	ਢ 2.49%	ਣ 4.87%	ਦ 3.91%	ਫ 2.77%	ਵ 1.78%		
ਸ	ਸ 75.19%	ਜ 1.80%	ਧ 7.48%	ਪ 1.88%	ਮ 3.22%	ਯ 4.29%	ਸ਼ 6.14%			
ਹ	ਹ 76.67%	ਚ 8.92%	ਠ 1.78%	ਰ 12.63%						
ਕ	ਕ 74.76%	ਟ 3.69%	ਢ 4.23%	ਦ 14.56%	ਫ 2.76%					
ਖ	ਖ 74.20%	ਘ 1.84%	ਥ 2.68%	ਧ 2.77%	ਪ 8.14%	ਬ 1.88%	ਮ 2.17%	ਖ਼ 6.32%		
ਗ	ਗ 80.14%	ਸ 3.92%	ਹ 0.45%	ਚ 0.90%	ਰ 2.34%	ਗ਼ 7.66%	ਸ਼ 4.59%			
ਘ	ਘ 80.32%	ਅ 14.33%	ਖ 5.35%							
ਚ	ਚ 74.23%	ਕ 2.14%	ਢ 6.89%	ਦ 6.41%	ਰ 10.33%					
ਛ	ਛ 79.46%	ਡ 2.94%	ਢ 4.53%	ਫ 4.48%	ਵ 8.59%					

Character	**Confused with characters**									
ਜ	ਜ 75.38%	ਸ 4.73%	ਸ਼ 4.42%	ਜ਼ 15.47%						
ਝ	ਝ 85.50%	ੲ 2.50%	ਕ 4.50%	ੜ 7.50%						
ਟ	ਟ 74.75%	ੲ 4.75%	ਣ 5.75%	ਦ 9.75%	ਫ 3.75%	ਫ਼ 1.25%				
ਠ	ਠ 76.00%	ਚ 2.00%	ਨ 12.00%	ਰ 2.00%	ਲ 8.00%					
ਡ	ਡ 70.00%	ਤ 12.00%	ਭ 13.00%	ੜ 5.00%						
ਢ	ਢ 68.00%	ੲ 4.00%	ਟ 5.00%	ਣ 4.00%	ਦ 7.00%	ਫ 12.00%				
ਣ	ਣ 72.54%	ੲ 4.86%	ਟ 9.84%	ਢ 3.42%	ਦ 2.89%	ਫ 2.67%	ਵ 3.78%			
ਤ	ਤ 70.84%	ਡ 6.39%	ਭ 12.66%	ੜ 10.11%						
ਥ	ਥ 72.26%	ਸ 1.98%	ਖ 2.77%	ਧ 8.23%	ਪ 3.74%	ਬ 8.92%	ਖ਼ 2.10%			
ਦ	ਦ 78.12%	ੲ 2.38%	ਚ 6.96%	ਟ 7.86%	ਢ 4.68%					
ਧ	ਧ 69.34%	ਸ 4.33%	ਖ 1.89%	ਘ 0.78%	ਥ 4.27%	ਪ 5.89%	ਮ 2.34%	ਯ 6.33%	ਸ਼ 2.86%	ਖ਼ 1.97%
ਨ	ਨ 82.14%	ਟ 1.78%	ਠ 9.31%	ਰ 0.76%	ਲ 3.87%	ਲ਼ 2.14%				
ਪ	ਪ 75.86%	ਸ 0.74%	ਖ 5.41%	ਘ 0.42%	ਥ 1.87%	ਧ 2.98%	ਬ 1.20%	ਮ 8.31%	ਯ 1.45%	ਖ਼ 1.76%
ਫ	ਫ 68.22%	ਛ 6.42%	ਟ 4.48%	ਢ 5.26%	ਣ 2.11%	ਦ 1.86%	ਵ 0.87%	ਫ਼ 10.78%		
ਬ	ਬ 77.64%	ਖ 4.36%	ਥ 13.39%	ਧ 2.94%	ਪ 1.67%					
ਭ	ਭ 75.89%	ਡ 7.71%	ਤ 12.12%	ੜ 4.28%						
ਮ	ਮ 78.48%	ਸ 4.81%	ਖ 3.41%	ਪ 6.88%	ਯ 1.06%	ਸ਼ 2.08%	ਖ਼ 3.28%			
ਯ	ਯ 78.11%	ਸ 2.42%	ਖ 0.86%	ਘ 2.86%	ਥ 4.51%	ਧ 6.02%	ਪ 1.85%	ਮ 1.14%	ਸ਼ 0.86%	ਖ਼ 1.37%
ਰ	ਰ 76.24%	ਹ 7.32%	ਚ 8.12%	ਟ 2.88%	ਠ 1.40%	ਦ 1.60%	ਨ 2.44%			
ਲ	ਲ 82.64%	ਸ 3.28%	ਨ 4.28%	ਸ਼ 2.46%	ਲ਼ 7.34%					
ਵ	ਵ 78.92%	ੳ 3.51%	ੲ 3.42%	ਛ 5.22%	ਣ 8.93%					
ੜ	ੜ 72.00%	ਡ 8.50%	ਤ 12.50%	ਭ 7.00%						
ਸ਼	ਸ਼ 74.52%	ਸ 12.89%	ਜ 1.24%	ਥ 0.86%	ਧ 1.52%	ਪ 0.64%	ਮ 2.68%	ਯ 0.48%	ਖ਼ 3.42%	ਜ਼ 1.75%

REFERENCES

[1] Beatrice Lazzerini, Francesco Marcelloni, ”Fuzzy classification of handwritten characters**”,** 18th International Conference of the North American, pp. 566 – 570,1999

[2] M Hanmandlu, K R Murali Mohan, Sourav Chakraborty, “Fuzzy logic based handwritten character recognition”, Proceedings of International Conference on image Processing, vol.3, pp**.** 42 – 45, 2001

[3] Weipeng Zhang, Yuan Yan Tang, Yun Xue, ” Handwritten Character Recognition Using Combined Gradient and Wavelet Feature”, International Conference on Computational Intelligence and Security, vol. 1, pp. 662 – 667, 2006

[4] Gary, M.T. Man and Joe, C.H. Poon, ” A fuzzy-attributed graph approach to handwritten character recognition”, Second IEEE International Conference on Fuzzy Systems, vol.1, pp. 570 – 575, 1993

[5] N. Liolios, E. Kavallieratou, N. Fakotakis, G. Kokkinakis, “A new shape transformation approach to handwritten character recognition”, 16th International Conference on Pattern Recognition , vol.1, pp. 584 – 587, 2002

[6] J. T. Tou and R. C. Gonzalez, “Recognition of Handwritten Characters by Topological Feature Extraction and Multilevel Categorization”, Computers, IEEE Transactions**,** Volume C-21, pp. 776 – 785, 1972

[7] L. Heutte, J.V. Moreau, T. Paquet, Y. Lecourtier and C. Olivier, “Combining structural and statistical features for the recognition of handwritten characters”, Proceedings of the 13th International Conference on Pattern Recognition**,** vol.2, pp. 210 – 214, 1996

[8] Nozomu Araki, Miyuki Okuzaki , Yasuo Konishi and Hiroyuki Ishigaki, ”A Statistical Approach for Handwritten Character Recognition Using Bayesian Filter”, 3rd International Conference on Innovative Computing Information and Control, pp**.** 194 – 194, 2008

[9] Jianming Hu; Hong Yan, ”Structural decomposition and description of printed and handwritten characters”, Proceedings of 13th International Conference on Pattern Recognition, vol.3, pp.230-234, 1996

[10] Tian Ming, Zhuang Yi, Chen Songcan, **“**Improving support vector machine classifier by combining it with k nearest neighbor principle based on the best distance measurement”, Proceedings. IEEE in Intelligent Transportation System, vol.1, pp**.** 373 – 378, 2003.

[11] Cao, L.J.; Chong, W.K,”Feature extraction in support vector machine: a comparison of PCA, XPCA and ICA”, Proceedings of the 9th International Conference on Neural Information Processing”, vol. 2, pp**.** 1001 – 1005, 2002

[12] Christopher J.C. Burges, “A Tutorial on Support Vector Machines for Pattern Recognition”, pp.1-43.