# Discrimination between Printed and Handwritten Gurmukhi Characters using Labelling Technique

Jaswinder Kaur

Research Scholar, Dept. of Computer Application, Gurukashi University, Talwandi Sabo, Punjab, India

**ABSTRACT:** Printed & handwritten text recognition is the part of pattern recoginition.In modern era, it has become essential to collect database for future use. This data may handwritten & machine written. In current work discrimination has performed with the help of three parameters like as PSNR, MSE, SSIM. Near about 100 images have taken for experiment & accuracy result came better than my previous research It also includes two main actions, first is preprocessing of image and other is segmentation.

**KEYWORDS:** PSNR, MSE, SSIM

## I. INTRODUCTION

The focus of current research is to discriminate printed and handwritten entries in data entry forms written in Gurmukhi language. Data entry forms are too common in all kinds of enterprises to collect customer information. This information is filled by customers manually in English or Arabic handwriting. But in these days, Gurmukhi has become part of popular language and it has become important to solve the issues which are related with Gurmukhi language. Data collection may be in form of Printed & handwritten.

Handwritten/machine-printed classification (HMC) is the process of labeling an image containing text segments, in order to discriminate handwritten from machine-printed text. It has numerous applications, particularly in (improving) Intelligent Character Recognition, automatic document analysis and anonymization [1]. OCR works with recognition of text.

**a) OCR (optical character reader)**
It is the mechanical or electronic conversion of images of typed, handwritten or printed text into machine-encoded text, whether from a scanned document, a photo of a document, a scene-photo or from subtitle text superimposed on an image.
There are number of difference between hand written and machine written text. These are given below:

- In machine, written language has less no. of strokes as compared to hand written text.
- Machine written text is more readable than the hand-written text.
- Machine written language has fixed sized text compare to the hand-written text.
- Machine written text has more options for formats compare to the hand-written text.
- Machine written text is equally spaced compare to the hand-written text.

**b) Gurmukhi Script**
Gurmukhi is the name of the script used in writing primarily Punjabi and, secondarily, in the Sindhi language. The word Gurmukhi seems to have gained currency from the use of these letters to record the sayings coming from the mukh (lit. mouth or lips) of the (Sikh) Gurus. The word Gurmukhi is the compound form of Guru and Mukh. However

the word Gurmukhi has been correctly used and understood for the 'holy utterances' from the Guru Granth Sahib. The letters no doubt existed before the time of Guru Angad Dev Ji (even of Guru Nanak) as they had their origin in the Brahmi, but the origin of the script is attributed to Guru Angad Dev Ji . He not only modified and rearranged certain letters but also shaped them into a script. He gave new shape and new order to the alphabet and made it precise and accurate. He fixed one letter for each of the Punjabi phonemes; use of vowel-symbols was made obligatory, the letters meant for conjuncts were not adopted and only those letters were retained which depicted sounds of the then spoken language. There was some rearrangement of the letters also such as sassa and hahaa were shifted to the first line and oorha was given the first place in the new alphabet.



Fig.1. Representation of Gurmukhi Script

## II. PROPOSED APPROACH

For the distinction between machine printed and handwritten text, we have to do following steps,fig.2
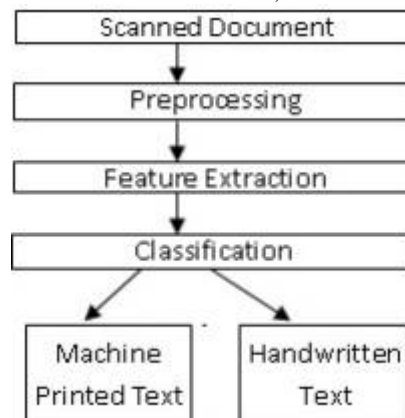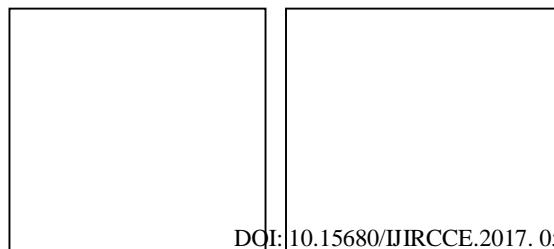


Fig.2. Block Diagram of System

To digitize the image I use iPhone with 8-megapixel camera. Here Figure 3,4 are images which are taken as input in different panels of Matlab GUI.

Input Images
Fig.3. Handwritten Text     Fig.4 Machine written Text

### A. PREPROCESSING

After the scanning process Image may contain a certain amount of noise. Depending on the resolution on the camera/ scanner and the success of the applied technique for thresholding, the characters may be broken and fade. Some of these defects, which may later cause poor recognition rates. To remove this problem preprocessing is performed. In addition to smoothing, preprocessing usually includes one more important feature to sharp the image is normalization. The normalization is a process which is applied for resizing, slant and rotation. Preprocessing has the following:

- Binarization
- Noise Removal
- Segmentation

- **Binarization**

A Binary Image is a digital image that has two possible values for each pixel. It has two Colors Black and White. Fig.5 shows the binarization of an image.

Fig.4. RGB Image, Greyscale Image, Binary Image

- **Noise Removal**

Erosion and dilation are the used for noise removal. These are the basic morphological tools for smoothening the binary image.

- **Segmentation**

Input image is first segmented into lines. A line is segmented on the basis of vertical gap between two lines. After a text line is segmented, it is scanned vertically and word is found by horizontal gap between two words.

Fig.5.  Line segmentation of Image

### B. FEATURE EXTRACTION

Feature extraction is used to extract relevant features for recognition of characters based on these features. First features are computed and extracted and then most relevant features are selected to construct feature vector which is used eventually for recognition. The computation of features is based on main two types of feature extraction: first is Statistical feature extraction & second is Structural feature extraction.

- **Statistical feature extraction:** In this type of extraction the extracted feature vector is the combination of all the features extracted from each character. The associated feature in feature vector of this type of extraction is due to the relative positions of features in character image matrix.

Fig.6. Example of Statistical feature

- **Structural feature extraction:** This is a primitive method of feature extraction which extracts morphological features of a character from image matrix. It takes into account the edges, curvature, regions, etc. This method extracts the features of the way character are written on image matrix.

Baseline

Lower Baseline

Fig.7.  Example of Structural feature

The functions that are used in feature extraction are:

    a) **Indexing and labelling:** This is a process by which distinct characters in an image are indexed and labelled in an image. Thus helps in classification of characters in image and makes feature extraction of characters simple.

Fig.8. Labelling of Image with green box

    b) **Boxing and Cropping:** This is a process of creating a boundary around the characters identified in an image. This helps by making cropping of characters easier. After boxing the characters are cropped out for storing them as input variables for recognition.

    c) **Reshaping and Resizing**: Reshaping is done to change the dimensions of the acquired character in desired shape. Resizing is done to reduce the size of characters to a particular minimum level.

## I.     PROPOSDED ALGORITHM

The classification is the process of identifying each character and assigning it to the correct character class. The approaches follow the labelling technique with parameters. The algorithm is as following :

Firstly the image is scanned, then image is preprocessed. Accordingly, image is threshold, denoised and then establish classifier rule for hand written and machine written text.

**Step 1** Input the images into the two different panels from any of the outer source.
**Step 2** Convert the image in to the matrix of rows and columns.
**Step 3** Convert the RGB image to Gray image.
**Step 4** Identify the threshold of the image based on OTSU technique.
**Step 5** Identify the Labels of the thresholded image.
**Step 6** Extract the text character based on labels of the image.
**Step 7** Segmented the text characters from the image.
**Step 8** Find out the value of image with different parameters like as PSNR, MSE, SSIM
**Step 9** Compare the Parameters & find out quality of machine written & handwritten character.

## II.    SIMULATION RESULTS

The simulation studies involves the a number of techniques which allow to extract the results with parameters. The proposed algorithm is implemented in MATLAB.We put the many data images to observe the result.
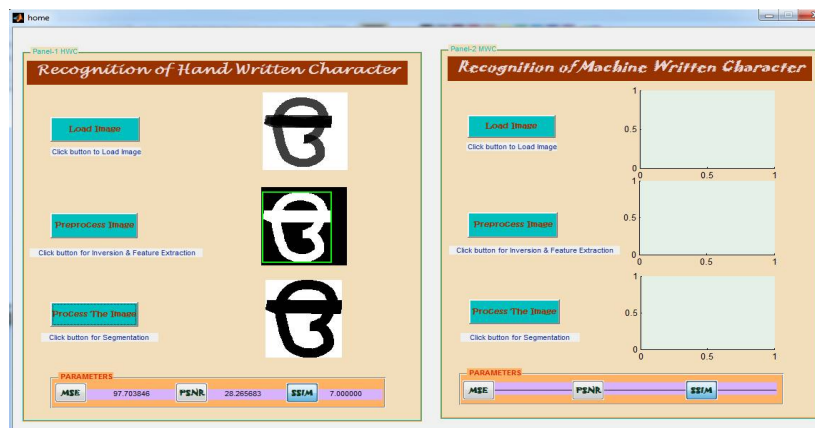


Fig.9. Example of input data  via MATLAB GUI

Result of test on 6 images is shown here in Table 1 & 2

| Handwritten Gurmukhi Character | Intensity Mean Value | MSE | PSNR |
|---|---|---|---|
| ੳ | 182.5572 | 97.703846 | 28.265683 |
| ਅ | 226.5384 | 64.523077 | 30.067649 |
| ੲ | 222.7957 | 64.369231 | 30.078016 |
| ਸ | 207.1419 | 68.150000 | 29.830141 |
| ਹ | 207.1419 | 70.857692 | 29.660929 |
| ਕ | 222.2701 | 68.796154 | 29.789158 |

Table.1 Parameters' presentation of Handwritten Gurmukhi character

| Machine Written Character | Intensity Mean Value | MSE | PSNR |
|---|---|---|---|
| ੳ | 200.7040 | 73.369231 | 29.509660 |
| ਅ | 222.6935 | 45.930769 | 31.543762 |
| ੲ | 214.8203 | 52.838462 | 30.935298 |
| ਸ | 222.6348 | 52.188462 | 30.989054 |
| ਹ | 218.7556 | 60.015385 | 30.382173 |
| ਕ | 222.9470 | 64.419231 | 30.074644 |

Table.2  Parameters' presentation of Machine written Gurmukhi character

As the two tables shows that Machine written character s(MWC) of Gurmukhi are  more clear as well as  MWC have less MSE ( Mean Squared Error). Due to less error, these characters have high PSNR as compared to handwritten characters.

## III.     CONCLUSION & FUTURE WORK

The simulation results showed that the proposed algorithm performs better with the machine written characters of Gurmukhi because handwritten characters are less clear than MWC.In future work, I will try to add  more parameters to represent the difference,

## REFERENCES

[1] Tanzila Saba., and A Nikolaidis.,"Language Independent Rule Based Classification of Printed & Handwritten Text". Proceedings of the IEEE 2015 tenth workshop on Multimedia Signal Processing, pp.393-398, 2015.
[2] B. B. Chaudhuri.,"Automatic Separation of Machine-Printed and Hand-Written Text Lines ", Pattern Recognition Letters ,2015.
[3] Surabhi Narayan., and A Nikolaidis.,"Discrimination of handwritten and machine Printed text is Scanner document Images based on Rough Set Theory". Proceedings of the IEEE 2012 tenth workshop on Multimedia Signal Processing, pp.393-398, 2012.
[4] Mrs.Saniya Ansari.,"Optimized and Efficient Feature Extraction Method for Devanagari Handwritten  Character Recognition", Pattern Recognition Letters ,2015.
[5] U.Pal., "Machine-printed and hand-written text lines identification". Proceedings of the IEEE 2001.
[6] Ergina Kavallieratou., and Stathis Stamatatos., "Discrimination of Machine-Printed from Handwritten Text Using Simple Structural Characteristics". Proceedings of the IEEE 2004.
[7] Purnendu Banerjee., and A Nikolaidis. ,"A System for Hand-Written and Machine-Printed Text Separation in Bangla Document Images". Proceedings of the IEEE 2012
[8] Konstantinos Zagoris., and Ioannis Pratikakis., "Automatic Classification of Handwritten and Printed Text in ICR Boxes". Proceedings of the IEEE 2014.
[9] Abhishek Jindal., and Mohd Amir., "Language Independent Rule Based Classification of Printed & Handwritten Text". Proceedings of the IEEE 2015 tenth workshop on Multimedia Signal Processing, pp.393-398, 2015.
[10] Ranjeet Srivastava., and Ravi Kumar Tewari., "Separation of Machine Printed and Handwritten Text for Hindi Documents ". Proceedings of the IEEE 2015
[11] Lincoln Faria da Silva., and Angel Sanchez.,"Automatic discrimination between printed and
handwritten text in documents". Proceedings of the IEEE 2009 .
[12] A. Saïdani .,and A. Kacem Echi., "Identification of Machine-printed and Handwritten Words in Arabic and Latin Scripts". Proceedings of the IEEE 2013

## BIOGRAPHY

Jaswinder  Kaur  is  a  Research  Scholar  of  M.Phil  in  Computer  Application,  UCCA College,Gurukashi University, Talwandi Sabo.She received master of Computer Application (M.Sc.) degree. Her research interests are computer networks, Pattern recognition,Web Designing,Image Processing etc.