www.ijecs.in
International Journal Of Engineering And Computer Science ISSN: 2319-7242
Volume 8 Issue 4 April 2019, Page No. 24603-24611
Index Copernicus Value (2015): 58.10, 76.25 (2016) DOI: 10.18535/ijecs/v8i04.4298

Classification of Printed and Handwritten Text Using Hybrid Techniques for Gurumukhi Script

Manpreet Kaur*[1], Balwinder Singh*[2]

[1]M.Tech. (Computer Engineering),
Yadavindra College of Engineering, Talwandi Sabo, India
[2]Assistant Professor (Computer Science),
Yadavindra College of Engineering, Talwandi Sabo, India

Abstract

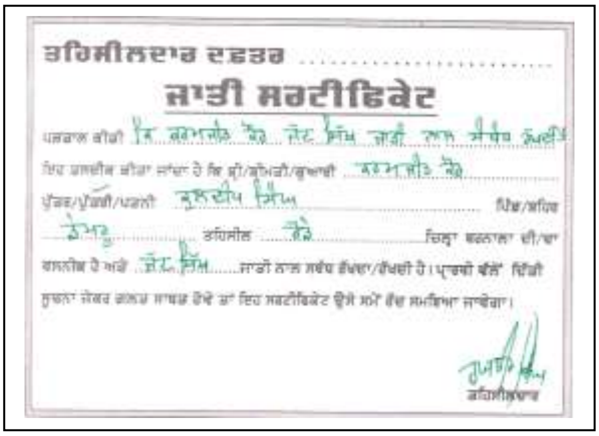
Text classification is a crucial step for optical character recognition. The output of the scanner is non-editable. Though one cannot make any change in scanned text image, if required. Thus, this provides the feed for the theory of optical character recognition. Optical Character Recognition (OCR) is the process of converting scanned images of machine printed or handwritten text into a computer readable format. The process of OCR involves several steps including pre-processing after image acquisition, segmentation, feature extraction, and classification. The incorrect classification is like a garbage in and garbage out. Existing methods focuses only upon the classification of unmixed characters in Arab, English, Latin, Farsi, Bangla, and Devnagari script. The Hybrid Techniques is solving the mixed (Machine printed and handwritten) character classification problem. Classification is carried out on different kind of daily use forms like as self declaration forms, admission forms, verification forms, university forms, certificates, banking forms, dairy forms, Punjab govt forms etc. The proposed technique is capable to classify the handwritten and machine printed text written in Gurumukhi script in mixed text. The proposed technique has been tested on 150 different kinds of forms in Gurumukhi and Roman scripts. The proposed techniques achieve 93% accuracy on mixed character form and 96% accuracy achieves on unmixed character forms. The overall accuracy of the proposed technique is 94.5%.

Keywords: Character classification, Mixed Character forms, Hybrid Techniques, Handwritten Gurumukhi Script.

1. Introduction

Optical character recognition and document image analysis are two curious topics in the field of pattern recognition. The aim of Optical Character Recognition (OCR) is to separate optical patterns (often contained in a digital image) corresponding to alphanumeric or other characters. Transmission and storage of information is done not only through computers but also through paper documents. To integrate these two mediums of information flow, a solution is for computer to “read” paper documents. Machine simulation of human reading is one of the areas, which has been the subject of research for the last three decades, yet it is still far from the final frontier. So, works are still going on this direction.

Each pattern having feature vector is classified in predefined classes using classifiers. Classifiers are first trained by a training set of pattern samples to prepare a model which is later used to recognize the test samples. The training data should includes of huge varieties of samples to recognize all possible samples during testing process. Some examples of generally practiced classifiers are Support Vector Machine (SVM), K- Nearest Neighbor (K-NN) and Probabilistic Neural Network (PNN).



ਤਹਿਸੀਲਦਾਰ ਦਫ਼ਤਰ ....

ਜਾਤੀ ਸਰਟੀਫਿਕੇਟ

ਪੜਤਾਲ ਕੀਤੀ ਕਿ ਕਰਮਜੀਤ ਕੌਰ ਜੱਟ ਸਿੱਖ ਜਾਤੀ ਨਾਲ ਸੰਬੰਧ ਰੱਖਦੀ

ਇਹ ਤਸਦੀਕ ਕੀਤਾ ਜਾਂਦਾ ਹੈ ਕਿ ਸ੍ਰੀ/ਸ੍ਰੀਮਤੀ/ਕੁਮਾਰੀ ਕਰਮਜੀਤ ਕੌਰ

ਪੁੱਤਰ/ਪੁੱਤਰੀ/ਪਤਨੀ ਕੁਲਦੀਪ ਸਿੰਘ .... ਪਿੰਡ/ਸ਼ਹਿਰ

[illegible] .... ਤਹਿਸੀਲ [illegible] .... ਜ਼ਿਲ੍ਹਾ ਬਰਨਾਲਾ ਦੀ/ਦਾ

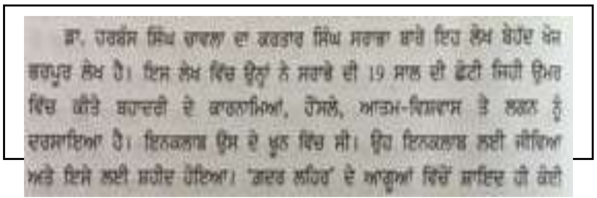
ਵਸਨੀਕ ਹੈ ਅਤੇ ਜੱਟ ਸਿੱਖ ......ਜਾਤੀ ਨਾਲ ਸਬੰਧ ਰੱਖਦਾ/ਰੱਖਦੀ ਹੈ। ਪ੍ਰਾਰਥੀ ਵੱਲੋਂ ਦਿੱਤੀ

ਸੂਚਨਾ ਜੇਕਰ ਗਲਤ ਸਾਬਤ ਹੋਵੇ ਤਾਂ ਇਹ ਸਰਟੀਫਿਕੇਟ ਉਸੇ ਸਮੇਂ ਰੱਦ ਸਮਝਿਆ ਜਾਵੇਗਾ।

ਤਹਿਸੀਲਦਾਰ

Fig.1.1 Handwritten and Machine printed text (mixed character) form

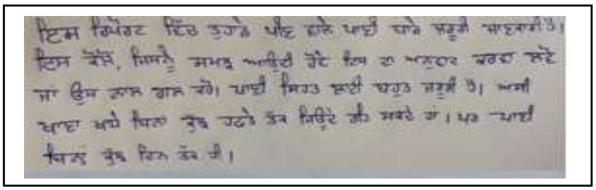
MACHINE-PRINTED text that are printed by a machine like as machine printed books, newspapers, magazines, Printed reports, printed certificates, printed applications , documents and various writing units in any video or any image. Machine printed characters are written in proper format like equal height, width, font style, font size for the all characters of any given document.



ਡਾ. ਹਰਬੰਸ ਸਿੰਘ ਚਾਵਲਾ ਦਾ ਕਰਤਾਰ ਸਿੰਘ ਸਰਾਭਾ ਬਾਰੇ ਇਹ ਲੇਖ ਬੇਹੱਦ ਖੋਜ ਭਰਪੂਰ ਲੇਖ ਹੈ। ਇਸ ਲੇਖ ਵਿਚ ਉਨ੍ਹਾਂ ਨੇ ਸਰਾਭੇ ਦੀ 19 ਸਾਲ ਦੀ ਛੋਟੀ ਜਿਹੀ ਉਮਰ ਵਿਚ ਕੀਤੇ ਬਹਾਦਰੀ ਦੇ ਕਾਰਨਾਮਿਆਂ, ਹੌਸਲੇ, ਆਤਮ-ਵਿਸ਼ਵਾਸ ਤੇ ਲਗਨ ਨੂੰ ਦਰਸਾਇਆ ਹੈ। ਇਨਕਲਾਬ ਉਸ ਦੇ ਖੂਨ ਵਿਚ ਸੀ। ਉਹ ਇਨਕਲਾਬ ਲਈ ਜੀਵਿਆ ਅਤੇ ਇਸੇ ਲਈ ਸ਼ਹੀਦ ਹੋਇਆ। 'ਗ਼ਦਰ ਲਹਿਰ' ਦੇ ਆਗੂਆਂ ਵਿਚੋਂ ਸ਼ਾਇਦ ਹੀ ਕੋਈ

Fig.1.2. Machine Printed Characters

Handwritten text can be divided into two categories: cursive and hand printed script. In this the characters are not in uniform and it can be change their size and style at the writing time according to every person even if characters are written by a same person then sometimes variation of characters are noticed.



ਇਸ ਰਿਪੋਰਟ ਵਿੱਚ ਤੁਹਾਨੂੰ ਪੀਣ ਵਾਲੇ ਪਾਣੀ ਬਾਰੇ ਜਰੂਰੀ ਜਾਣਕਾਰੀ ਹੈ। ਇਸ ਬਾਰੇ, ਕਿਸੇ ਨੂੰ ਸਮਝ ਆਉਂਦੀ ਹੋਵੇ ਇਸ ਦਾ ਅਨੁਵਾਦ ਕਰਵਾ ਲਵੋ ਜਾਂ ਉਸ ਨਾਲ ਗੱਲ ਕਰੋ। ਪਾਣੀ ਸਿਹਤ ਲਈ ਬਹੁਤ ਜਰੂਰੀ ਹੈ। ਅਸੀਂ ਖਾਣਾ ਖਾਧੇ ਬਿਨਾ ਕੁਝ ਹਫਤੇ ਤੱਕ ਜਿਉਂਦੇ ਰਹਿ ਸਕਦੇ ਹਾਂ। ਪਰ ਪਾਣੀ ਬਿਨਾ ਕੁਝ ਦਿਨ ਤੱਕ ਹੀ।

Fig.1.3. Handwritten Characters

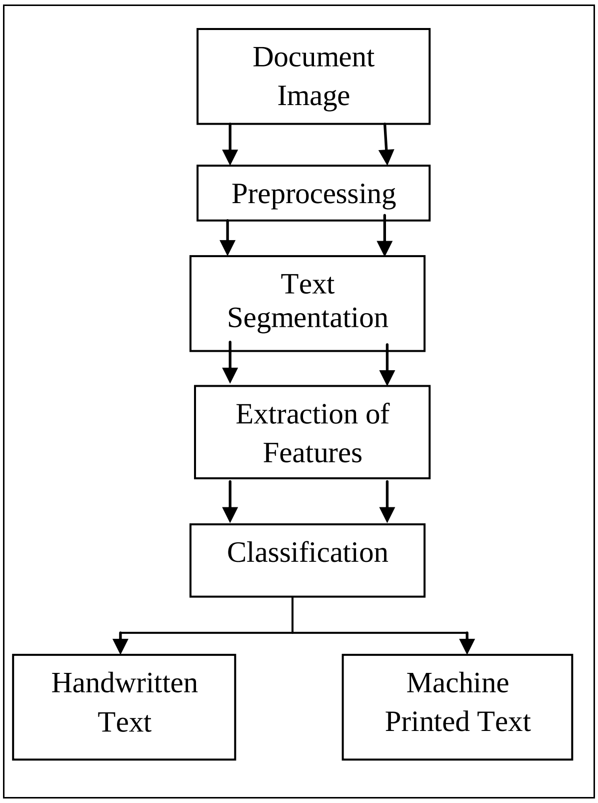
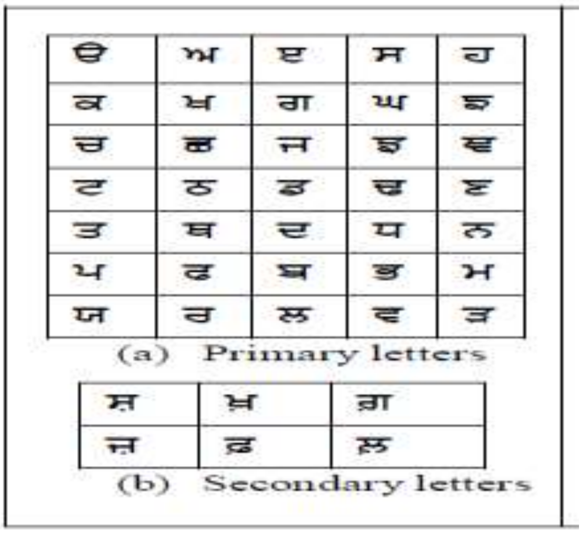


Fig.1.4. Overview of the system [11]

1.1 Overview Of Gurumukhi Script

Gurumukhi is derived from the combination of two words "Guru" and "Mukh". Gurumukhi means to record the sayings from the mukh of the Gurus, i.e. from the Guru's mukh. The credit to originate this script goes to Guru Angad Dev Ji. Gurumukhi script consists of 35 primary letters and 6 secondary letters. The first three letters are made to represent ten vowel sounds; last five letters are called semi vowels and 3 half characters which lie at the feet of consonants. The character set of Gurumukhi script is shown in table 1.1, 1.2, 1.3, 1.4 respectively.

Table 1.1 Character Set of Gurumukhi Script



ੳ	ਅ	ੲ	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

(a) Primary letters

ਸ਼	ਖ਼	ਗ਼
ਜ਼	ਫ਼	ਲ਼

(b) Secondary letters

Table 1.2: Vowels and corresponding modifiers

ਅ	ਆ(◌ਾ)	ਇ (ਿ◌)
ਈ (◌ੀ)	ਉ (◌ੁ)	ਊ (◌ੂ)
ਏ (◌ੇ)	ਐ (◌ੈ)	ਓ (◌ੋ)
ਔ (◌ੌ)		

Subjoined Consonants (Half Characters)

Some Punjabi words require consonants to be written in a conjunct form, which takes the shape of a subscript to the main letter. There are only three commonly used subjoined letters.

Table 1.3 Half Characters

◌੍ਰ	◌੍ਹ	◌੍ਵ

Other Symbols

Besides the consonants and the vowels, other constituent symbols in Gurumukhi are a set of vowels modifiers called *matra* placed to the left, right, above or at the bottom of a character or conjunct, pure consonants forms corresponding to some consonant (also called half letters) which when combined with other consonants yield conjuncts.

Table 1.4 Other Symbols

◌ੰ (tippi)	◌ਂ (bindi)
◌ੱ (adhak)	◌ਃ(visarg)

2. Literature Survey

Saba et. al (2015) [1]: In is paper used the new statistical and structural features of text lines to classify them into separate categories and this technique is independent of language, style, size and fonts. This approach achieved the 90% accuracy and this technique is tested on Arab and English language.

Srivastava et. al (2015) [2]: Statistical and structural features are used to distinguish between handwritten and machine printed of hindi documents. The basic advantage of this approach is that it is font independent and size independent for Hindi language not suitable for multilingual data and the overall accuracy of the system is 94.1%.

Jindal et. al (2014) [3]: This paper proposed a technique to classify the handwritten and machine printed characters inside the intelligent character recognition (ICR) cells. The overall accuracy is achieved 91%.

Saidani et. al (2014) [4]: This technique is tested on Arabic and Latin scripts. This paper uses a Pyramid Histogram of Oriented Gradients (PHOG) features. The PHOG features are counts occurrences of gradient orientation in localized portion of an image. Experiments have been conducted using standard databases and it achieved 98.3 % accuracy.

Saïdani et. al (2013) [6]: This novel approach is checked on Arabic and Latin scripts. This paper uses a new structural features for conclude their results. The tested data performed their separation with various font styles and sizes, written in Arabic and Latin scripts and correct classification of 98.4% for word level script and nature identification using Bayes classifier.

Narayan et. al (2012) [8]: The Rough Set theory is discrimination of handwritten and machine printed text is achieved by the consideration of uniform occurrence as a main feature. The experiments have been tested on locally generated 400 samples and their samples are collected from the IAM data set.

Banerjee et. al (2012) [9]: Though developed for Bangla script, this approach can be used for Devanagari script, because the headline and vertical lines are dominant shapes in Devanagari script as well. The overall accuracy of system is 96.49%.

Mozaffari et. al (2012) [10]: This approach is tested on the farsi language and Arabic language documents and firstly, finding the word blocks in given document, then three different features sets were extracted. They include two well established features, previously used for latin script handwritten from machine printed text separation, and a new feature, called baseline profile. SVM and KNN classifiers were utilized to separate handwritten and machine printed words with 97.1% accuracy.

Silva et. al (2009) [13]: In this authors discrimination is done by using the four steps: digitalization, preprocessing, feature extraction and decision or classification. In this the data mining technique is mainly used for the classification or decision step with 80% accuracy.

Zheng et. al (2004) [14]: This technique is a novel aspect for noisy documents that are written in Chinese language and we treat noise as a separate class and model noise based on selected features. In this for the classification of handwritten and machine printed text the Trained Fisher classifiers are used.

Kavallieratou et. al (2004) [15]: In this paper an integrated system is capable to localize text areas and split them in text-lines is used. In this the mainly structural characteristics are used that capture the differences between machine-printed and handwritten text-lines is introduced for Latin scripts.

Guo et. al (2001) [16]: This paper present an algorithm that is focused on the theory of hidden Markov models (HMM) to done their separation between machine printed and handwritten materials. This approach is well suitable for English texts and Latin scripts only with the training with 92.86 accuracy.

Pal et. al (2001) [17]: This method is concludes the results on the Bangla scripts and Devnagari scripts. In this paper the mainly structural and statistical features are used to obtain their results with 98.6% accuracy.

3. Existing Techniques

The various techniques used to separate the handwritten and machine printed character are given below:

Structural Features: The first type of features that are extracted from text is structural features. Accordingly, the baseline and lower baseline of the text is detected using enhanced horizontal histogram. Number of strokes is determined between baseline and lower baseline. It is found that for the printed text, there exist only few strokes below baseline due to presence of characters that have descenders. But, for handwritten text, there is an ample amount of strokes below baseline [1].

Statistical Features: Second type of extracted features is statistical features that analyze standard deviation of stroke thickness at the contours only. Accordingly, the machine printed text has stable stroke thickness at the contours and therefore, standard deviation is minimum whereas it varies in case of script writing [2].

Pyramid histogram of oriented gradient (PHOG): Pyramid HOG (PHOG) takes the spatial property of the local shape into account while representing an image by HOG. The spatial information is represented by tiling the image into regions at multiple resolutions based on spatial pyramid matching. Each image is divided into sequence of increasingly finer spatial grids by repeatedly doubling the number of divisions in each axis direction. In this the points in each grid is counted. In this the counted points of cells at one level the sum over that four cells and then in next level this becomes a pyramid representation [4].

SVM: SVM are supervised learning methods, which have been widely and successfully used for pattern recognition. The main concept of SVM lies to find a hyper plane that allows separating two classes, leaving the largest margin between the vectors of the two classes. However, in real life, problems can be linearly non separable. To deal with this problem, a nonlinear decision surface is obtained by lifting the feature space into a higher dimensional space. A linear separating hyper plane is found in the higher dimensional space that gives a nonlinear decision surface in the original feature space [9].

Random Transform: The Radom transform computes projections of an image along specified directions. A projection of a two-dimensional function I (x, y) is a set of line integrals. The Radom transform computes the line integrals from multiple sources along parallel paths in a certain direction. To represent an image, the Radom transform takes multiple and parallel projections of the image from different angles by rotating the source around the centre of the image [11].

3.1 Proposed Classification Techniques

(i) LAB: The **Lab** color space describes mathematically all perceivable colors in the three dimensions, **L** for lightness and **a** and **b** for the color opponents green–red and blue–yellow. The three coordinates of CIELAB represent the lightness of the color (**L*** = 0 yields black and **L*** = 100 indicates diffuse white; specula white may be higher), its position between red/magenta and green (**a***, negative values indicate green while positive values indicate magenta) and its position between yellow and blue (**b***, negative values indicate blue and positive values indicate yellow). This technique is used to solve the problem of mixed Gurumukhi characters on various forms [5].

(ii) K-means clustering: The K-means clustering is used for classification of object based on a set of features into K number of classes. The classification of object is done by minimizing the sum of the squares of the distance between the object and the corresponding cluster. [12]

The algorithm for K –means Clustering:

1. Pick center of K cluster, either randomly or based on some heuristic.
2. Assign each pixel in the image to the cluster that minimizes the distance between the pixel and the cluster center.
3. Again compute the cluster centers by averaging all of the pixels in the cluster. Repeat steps 2 and 3 until convergence is attained.

(iii) Otsu Threshold Algorithm: Thresholding creates binary images from grey-level images by setting all pixels below some threshold to zero and all pixels above that threshold to one. It was defined as:

i) According to the threshold, separate pixels into two clusters

ii) Then find the mean of each cluster.

iii) Square difference between the means.

iv) Multiply the number of pixels in one cluster times the number in the other.

3.2 Proposed Work

The proposed work includes the following steps:

Step-1 Scanning Image

In this step the document is converted into scanned image with the help of image scanner. In this step we give the scan image as a input to system. Scanning image is shown in figure 3.1.

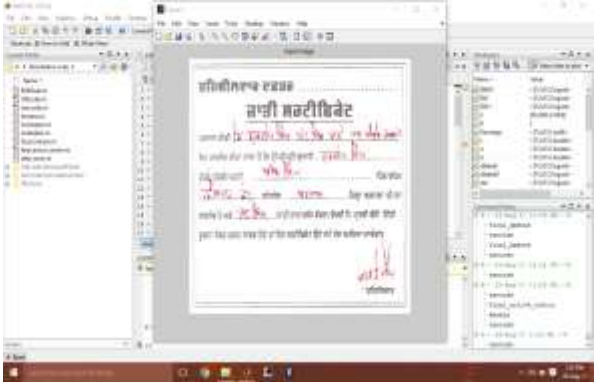


Figure 3.1: Handwritten and Machine printed form(Mixed Character form) written in Gurumukhi Script

Step-2 RGB Conversion

In this step the input images is converted into RGB image i.e. in the form of Red, Green and Blue. In this step the input image is split into three modules RED, GREEN and BLUE and then apply operation separately on each module for further working.

Step-3 Noise Removal

In this step the noise means extra pixels are removed from the image by adjusting image intensity or by imadjust command.

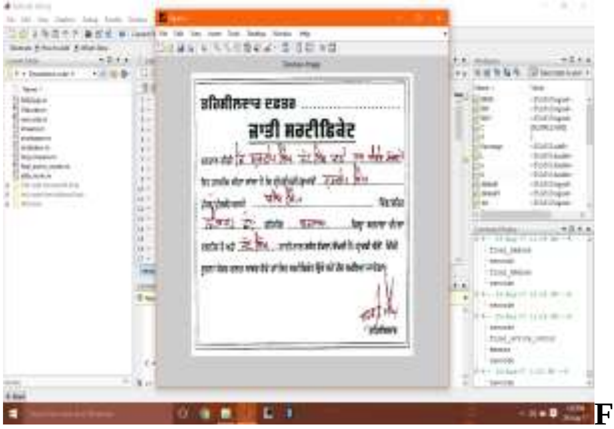


Figure 3.2: Noise Removal from input image

STEP-4 SKEW REMOVAL

Skew removal: In this step the skewness is removed by using imdeskew command.

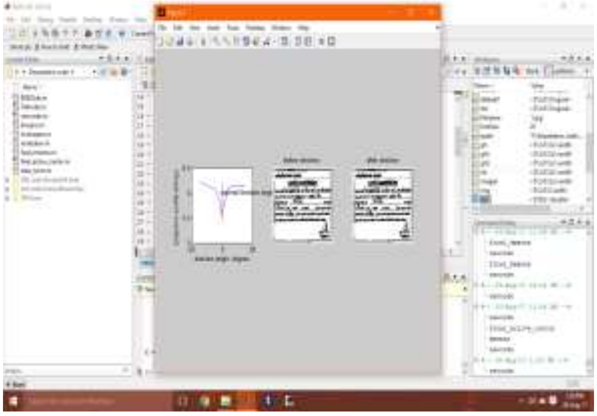


Figure 3.3: Skew removal

Step-5 Color Conversion

In this step, the R+G+B components are combined and converted into the color image.

Step-6 Lab Color Space Conversion

In this step, the color image is converted into LAB color space.

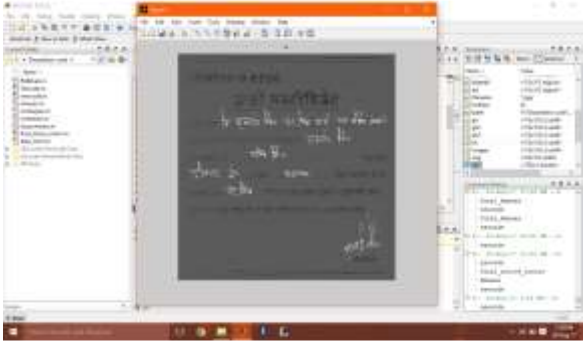


Figure 3.4: Color Conversion

Step-7 Apply Clustering

Apply Clustering for extracts the highlights and non highlights characters of LAB color space image.

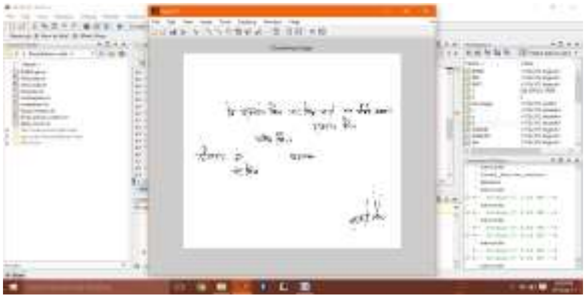


Figure 3.5: Apply Clustering on LAB color space image

Step-8 Binary Conversion

In this step, the LAB image is converted into the Binary image by using Otsu method for setting the thresholding value.

STEP-9 Classification Of Handwritten And Machine Printed Characters

In this step the handwritten and machine printed characters are classified.

1. Handwritten Character:- The Handwritten Characters are non uniform and can vary greatly in size and style. In the location of characters is not predictable, nor the spacing between them. In an unconstrained system, characters may be written anywhere on the page and may be overlapped or disjoint.

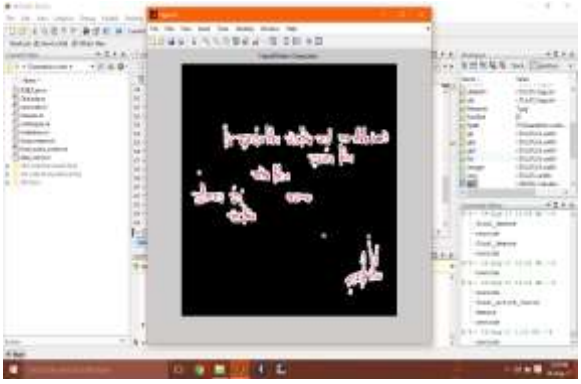


Figure 3.6: Handwritten Character

2.Machine Printed Character:- Machine printed characters are uniform in height, width and pitch assuming the same font and size are used. Problem related to these are solved with little constraint.

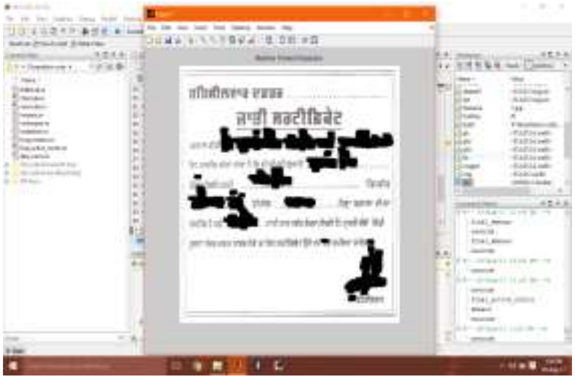
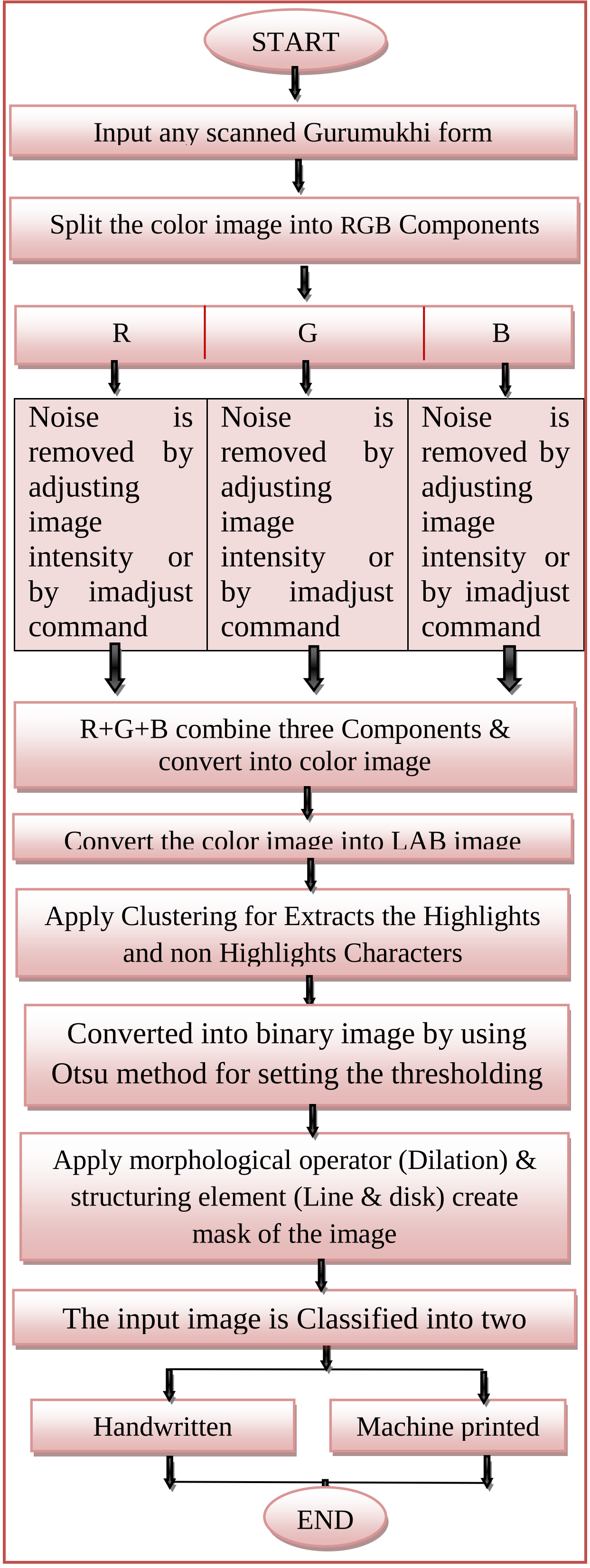


Figure 3.7: Machine Printed Character

3.3 Flow Chart For Proposed Work

The proposed flowchart is shows the input, processing and output in the pictorial form. The working of proposed technique in the form of steps is shown as below:


START
Input any scanned Gurumukhi form
Split the color image into RGB Components
R
G
B
Noise is removed by adjusting image intensity or by imadjust command
Noise is removed by adjusting image intensity or by imadjust command
Noise is removed by adjusting image intensity or by imadjust command
R+G+B combine three Components & convert into color image
Convert the color image into LAB image
Apply Clustering for Extracts the Highlights and non Highlights Characters
Converted into binary image by using Otsu method for setting the thresholding
Apply morphological operator (Dilation) & structuring element (Line & disk) create mask of the image
The input image is Classified into two
Handwritten
Machine printed
END

4. Result And Discussion

4.1 Representation Of Results

The proposed technique is tested on various kinds of the forms such as dairy farm, university forms, application forms, registration forms, self declaration form etc. The input and the output of the proposed technique are shown below:

Firstly, take the declaration form as a input shown in figure 4.1 and after performing various steps on this input and the generated output is shown in figure 4.2 and 4.3

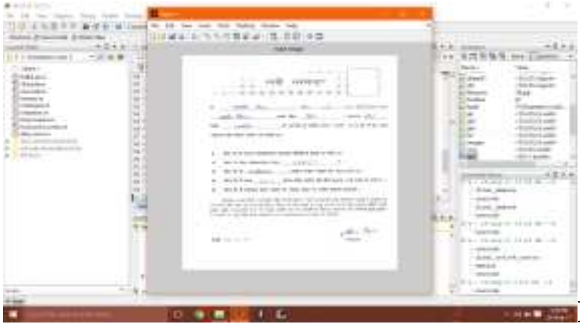


Figure 4.1: Input image

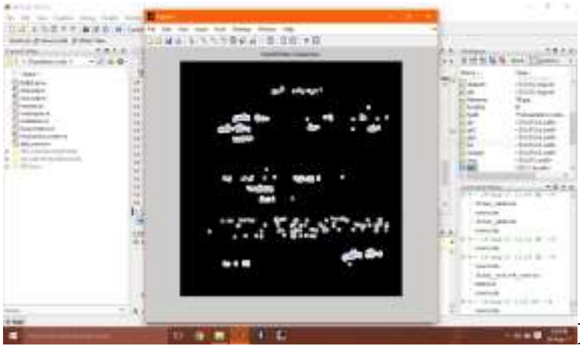


Figure 4.2: Handwritten Character

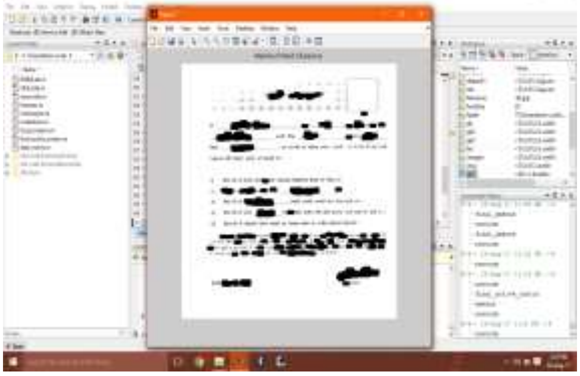


Figure 4.3: Machine Printed Character

4.2COMPARISON OF PROPOSED WORK WITH EXISTING WORK

Here, we compare the proposed work with the existing system. From comparison, it is clear that the proposed system is better than the existing system. Comparison is shown in table 4.1 and Table 4.2 shows the results obtained by our proposed system.

Table 4.1: Comparison with Existing technique

Factors	**Existing Work**	**Proposed Work**
Technique Used	Statistical and Structural Features[1]	Hybrid Techniques
Classification	Existing work does not Classify the Mixed characters(Printed and Handwritten Characters)	Proposed work solves the Mixed characters classification problems
Accuracy	90%	94.5%

Table 4.2: Different Phases of Words Showing Accuracy

Sr No.	**Test Case Type**	**No. of Test Cases**	**No. of Successful Test cases**	**Accuracy**
1.	Mixed character form	100	93	93%
2.	Words contain not mixed characters	50	48	96%

Here we have tested this algorithm on 150 handwritten words taken from different people with different handwriting. The overall accuracy for classification comes out to be 94.5% which is very good.

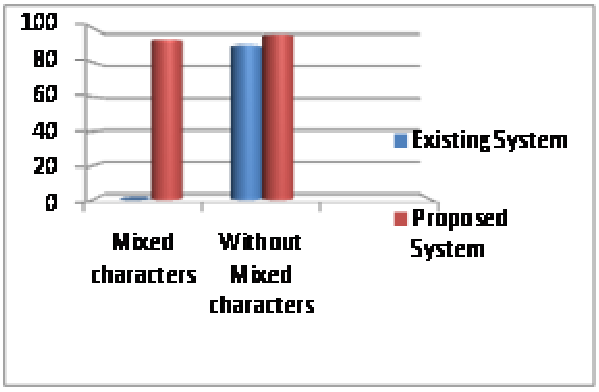


Figure 4.4: shows the graph representation of existing as well as proposed system.

5. Conclusion

Classification of machine printed and handwritten text is uphill task due to variations in handwriting, structural properties of language and presence of upper, lower modifiers. Though these challenges made classification very difficult but a new algorithm has required to be implemented to overcome these difficulties. Without an efficient approach, it becomes very complex job to classify the mixed character. In this research work a new approach for classification of various characters is to be implemented. The new approach has been tested on various documents containing different types of forms. Here we have tested this algorithm on 150 form images taken from different people with different handwriting style. The overall accuracy for classification of machine printed and handwritten text comes out to be 94.5% which is very good. In the proposed system a new technique is developed to classify mixed forms named as Hybrid Techniques. The existing techniques does not solve the mixed character classification problem but the proposed technique is solves this problem with better accuracy.

Table 5.1. Comparative study of existing work on different characters:

Reference	**Technique used**	**Type of script**	**Accuracy**
Saba et. al (2015) [1]	Structural and Statistical Features	English & Arab	90%
Mozaffari et. al (2012)[10]	SVM and KNN Classifier	Farsi and Arabic	97.1%
Banerjee et. al (2012)[9]	SVM	Bangla documents	96.49%
Zemomi et. al (2011) [11]	Random transform	IAM database	98%
Pal et. al (2001) [17]	Structural and Statistical Features	Bangla and Devnagri	98.6%

6. Future Work

The proposed system can be extended to increase of the proposed work by using more techniques. In future need to develop a technique that is capable to classify the handwritten and machine printed characters in noisy form documents.

References

1. Saba T., Almazyad A.S., Rehman A. "Language Independent Rule Based Classification of Printed & Handwritten Text", International conference on evolving and adaptive intelligent system (EAIS), pp.1-4 December 1, 2015.
2. Srivastava R.,Tewari R.K., Kant S., "Separation of Machine Printed and Handwritten Text for Hindi Documents" International research journal of engineering and technology(IRJET), Vol.2, Issue 2, pp.704-708, 2015.
3. Jindal A., Amir M., "Automatic Classification of handwritten & printed text in ICR Boxes", International advance Computing Conference(IACC), IEEE, pp.1028-1032, 21 Feb., 2014.
4. Saïdani A, and Echi A.K.. Belaid A., "pyramid histogram to oriented gradient for machine-printed/handwritten and Arabic/latin words discrimination", 6th international conference of soft computing and pattern recognition, IEEE, pp.267-272, 11 Aug., 2014.
5. Wang X., Hansch R., Ma L., Hellwich O., "Comparision of different Color Spaces for Image Segmentation Using Graph Cut", International Conference on Complex Vision theory and Applications(VISAPP), Vol.1, pp. 301-308, 5 Jan.,2014.
6. Saïdani A, and Echi A.K.. Belaid A. "Identification of Machine-printed and Handwritten Words in Arabic and Latin Scripts", 12th International conference on document analysis and recognition, IEEE, pp.798-802, 25 Aug., 2013.
7. Zagoris K.et. al, "Handwritten and Machine printed text separation in document images using the Bag of Visual

Words Paradigm", International Conference on Frontiers in Handwriting Recognition, IEEE, pp.103-108, 18 Sep., 2012.

8. Narayan S., Gowda S.D., "Discrimination of handwritten and machine Printed text is Scanner document Images based on Rough Set Theory" World Congress on Information and Communication Technologies, IEEE, pp.590-594, 30 Oct., 2012.
9. Banerjee P., Chaudhari B.B., "A System for Hand-Written and Machine-Printed Text Separation in Bangla Document Images" International Conference on Frontiers in Handwriting Recognition, IEEE, pp. 758-762, 18 Sep., 2012.
10. Mozaffari S., Bahar P., "Farsi/Arabic handwritten from machine printed words discrimination", International Conference on Frontiers in Handwriting Recognition, IEEE, pp. 698-703, 18 Sept., 2012.
11. Zemouri ET-T., Chibani Y., "Machine printed handwritten text discrimination using random transform and SVM classifier", 11th International Conference on Intelligent Systems Design and Applications, IEEE, pp.1306-1310, 22 Nov., 2011.
12. Sulaimen S. N., Isa N. A. M., "Adaptive Fuzzy –K-Means Clustering Algorithm for Image Segmentation", IEEE Transactions on consumer electronics, Vol.56, Issue 4, pp. 2661-2668, Nov 2010.
13. Silva et. al, "Automatic discrimination between printed and handwritten text in documents", Brazilian symposiam on computer graphics and image processing, IEEE, pp. 261-267, 11 Oct., 2009.
14. Zheng et. al, "Machine printed text and handwriting identification in noisy document images ", IEEE transaction on pattern analysis and machine intelligence, Vol 26, No 3, pp. 337-353, 26 Mar., 2004.
15. Kavallieratou E., Stamatates S., "Discrimination of Machine-Printed from Handwritten Text Using Simple Structural Characteristics", 17th international conference on pattern recognition (ICPR), IEEE , Vol.1, pp.437-440, 23 Aug., 2004.
16. Guo J. K. , Ma M. Y., "Separating handwritten material from machine printed text using hidden Markov Models", 6th international conference on analysis and recognition (ICAR), pp. 439-443,2001.
17. Pal U., Chaudhuri B.B., "Machine-printed and hand-written text lines identification", Pattern recognition letter, Vol.22, Issue 3, Elsevier, pp. 431-441, 31 Mar., 2001.