iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 **Issue:** XII **Month of publication:** December 2020

DOI: https://doi.org/10.22214/ijraset.2020.32614

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue XII Dec 2020- Available at www.ijraset.com

Analysis of Handwritten Joint Characters in Gujarati Language

Rachana Chaudhari[1], Purna Tanna[2]
[1, 2]*Assistant Professor, GLS University*

Abstract: ***Handwritten Character Recognition (HCR) has been challenging research especially in Gujarati language. Handwritten Character Recognition may be online and offline. Gujarati script contains numerical, characters and modifiers. In Gujarati character, when two consonant joints they form conjunct character or also called joint character. This adds more problems for identification of existing character. These are the some consonants of Gujarati script which can be considered as reason for low accuracy in Gujarati characters recognition. This paper describes an analysis of Gujarati handwritten joint characters.***

Keywords: OCR (Optical Character Recognition), HCR (Handwritten Character Recognition), Joint Character, Conjunct Consonants, Half Join Character, Vertically Join Character, Touching Character

I. INTRODUCTION

There are lots of documents that written in Gujarati language from which some of handwritten and some of printed. Optical Character Recognition (OCR) can help to convert that document from printed or written to editable format.

Optical Character Recognition (OCR) translate handwritten or printed scan image into machine editable text format. The advantage is that the textual material can be edited which is generated from scanned image. OCR is a popular area of research in current era of technology and lots of work has been done and research is still going on for better results. Handwritten Character Recognition (HCR) interprets user's character which may be source from paper document, image or from devices into a format that understands by machine. The source mode may be online or offline. In Gujarati language, developing accurate handwritten character recognition system becomes a huge challenge. Various techniques have been proposed towards Indian languages in handwritten character recognition but comparatively very less work has been reported for Gujarati language. The main reason for this is different people having different handwritten style and handwriting of same writer on a particular character may differ at various instances depending upon mood and environment so it is difficult to recognition. Handwritten character recognition helps in some application like printed or handwritten document reading, bank cheque processing, automatic data entry application, etc.

II. JOINT CONSONANTS OR CONJUNCT CONSONANTS

Gujarati script contains 13 vowels, 34 consonant and 10 numerical digits. Here, Character recognition can be of two types: Online Character Recognition and Offline Character Recognition.

Figure 1: Gujarati Consonants

Online character recognition which takes input at run time and record sequence of pen coordinates. Offline Character Recognition which takes scanned image as an input and after processing on it we can get characters from that. The Offline Character Recognition further divided into two types: Handwritten Character and Printed Character. Handwritten Character Recognition is more difficult compare to Printed Character because handwriting style of human. In printed character recognition we can get standard font style and size.

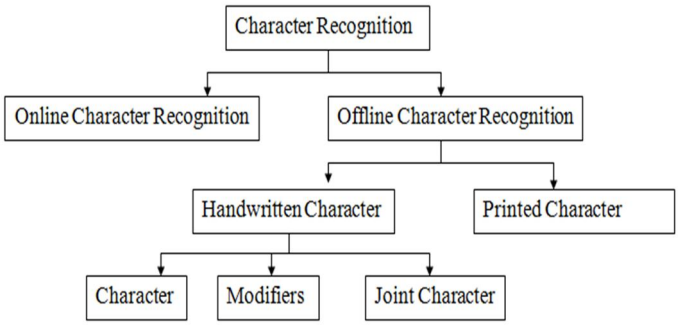


Figure 2: Character Recognition

In Gujarati words joint character and modifiers are separate entities. Consonant cluster or Joint character is a combination of two consonants. This combination adds more problem of identifying and recognition of combined consonants. Joint characters of Gujarati script is considered as a reason for low accuracy or less progress in analysis and recognition of Gujarati character.
There are three types of form for Joint Characters. In Joint characters, consonants can be joining as half join form, vertically joint form or touching form.

1) *Half Join Form:* As shown in below figure a half joint form is created by removing the vertical line in the consonant shape of first character. Here first character removes its half portion of character and joint with another character.

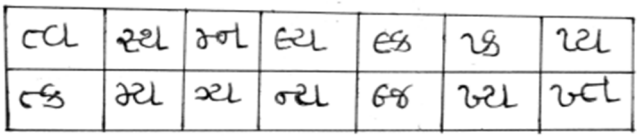


Figure 3: Half Join Form

2) *Vertically Join Form:* As shown in below figure a vertically join form is combination formed by subjoining two consonants vertically. Here first character which is join with another character vertically where second character remove its half portion from hole character.

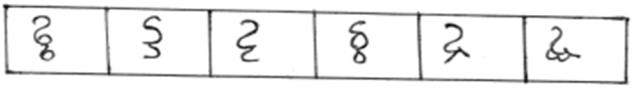


Figure 4: Vertically Join Form

3) *Touching Form:* As shown in below figure a touching form where character comes closer together without major shape changes. Here in this form two characters are joint in such way that they are just joint with each other but does not remove any portion from its own character.

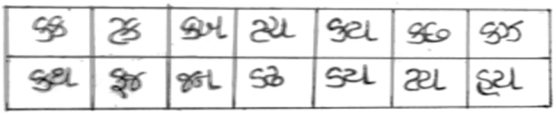


Figure 5: Touching Form

The prime theme of the study is to analyze characteristics of the Joint character, developing character recognition engine capable of identifying character from image and converting it to editable text format by suggesting suitable character recognition approach.

III. LITERATURE SURVEY

A. Arpit A. Jain and Harshal A. Arolkar [1] presented a survey on Gujarati Handwritten Character recognition techniques. They have differentiated Online Character and Handwritten Character. They have mentioned that recognizing similar looking character and conjunct character is little bit complex. In this survey they have presented that results for individual characters, numbers or sample are not providing accuracy of recognizing characters at satisfactory level.

B. Dr. Dinesh Satange, Dr. P E Ajmire and Fozia I Khandwani [2] presented the method for handwritten Gujarati numeral recognition using neural network. They have designed database for handwritten numeral with the help of different age group. They have used Support Vector Machine and Multi-Layer Perception for the classification. They have use MLP classifiers for training, cross validation and testing. They have got accuracy of recognition is 95.12% on training, 90% on Testing and 93.83% on cross validation.

C. V. A. Naik and A. A. Desai [3] have used hybrid feature set and compared SVM with linear, polynomial and radial basis function kernels for classification. They have used zoning and chain code directional feature. They proposed an algorithm for online handwritten Gujarati numeral recognition using feature set of 22 different structural and statistical features. They have tested on training set of 2000 samples collected from 200 different writers of different age group and each writer given 10 samples. They have achieved an accuracy of 92.60% for linear with processing time of 0.13 seconds, 95% for polynomial with processing time of 0.15 seconds and 93.80% for RBF kernel with processing time of 0.18 seconds per stroke.

D. Nikita Mehta and Jyotika Doshi [4] reviewed and analyzed various techniques for Handwritten Character Recognition. They have given overview of exist different approaches for Handwritten Character Recognition and define steps like Grayscale Conversion, Binarization, Noise Removal, Skew Detection and correction, Normalization, Segmentation and Feature Extraction and Classification for OCR. They have mentioned that character with different modifiers and conjunct characters are complex to identify and need lots of research for accurate result.

E. Sumit G. Trivedi and Arun Nandurbarkar [5] have divided system into three stages. In First stage for image acquisition and pre processing they have different four steps like Thresholding, Filtering, Boundary Tracing and Skeletonization. In Second stage for feature extraction they have used four techniques Correlation Function based Features like Invariant Moment, Projection Profile and Gradient Feature. In third stage of classification they have used three techniques namely Support Vector Machine, K Nearest Neighbour and Naïve Bayes Classifier. In system they have dataset contains 150 samples for each Gujarati Character that is total 5100. For classification they have used 10 fold cross validation and got accuracies 97.92% for SVM and 97.27% for K-NN using gradient feature and hybrid combination of feature.

F. Bhumika B. Patel and Hinaxi M. Patel [6] have defined main component of character recognition like preprocessing, segmentation, extraction, classification and post processing. They have concluded that many researchers achieve good recognition rate for individual character but character with different modifiers and conjunct characters are complex to identify. No such OCR available for Gujarati handwritten script with 100% accuracy. Researchers have been working with Gujarati Character Recognition but still it remains open problem in transforming the document into digitized form.

G. Nishu Rastogi, Maitreyee Dutta and Ajay Indian [7] want to find an efficient algorithm for feature extraction with minimum length of feature vector. They have used sample set of 2500 containing 250 sample of each digit from people of different age group and with image format jpg with resolution 2400 x 4800 dpi. A feature vector of length nine using directional gradient histogram (DGH), eight using chain coding (CCH) and thirty two using chain code by dividing image into four parts (QBCCH) is passed through the Backpropagation Neural Network with Levenberg-Marquartdt training function. With combinational feature vector using DGH with CCH and DGH with QBCCH resulted in much higher recognition rates for isolated Gujarati Numerals. Using combinational feature vector method DGH with QBCCH they have got 96.8% accuracy for recognition.

H. Milind Kumar Audichya and Jatinderkumar R. Saini [8] presented use of Tesseract to recognize Printed Gujarati characters with the help of available trained data set for Gujarati script. They have used newspaper, books and official documents as image with different font style, size and color and got more than 80% of accuracy for digitally typed printed images. They have defined how Tesseract works and can be used as an initial stage for recognition of character for future research work.

I. Madhuree Ardeshana, Ankit Sharma, Dipak Adhyaru and Tanish Zaveri [9] used Discrete Cosine Transform (DCT) method for feature extraction and Naive Bayes (NB) method for classification of handwritten Gujarati Character. In this paper they have shown architecture of OCR. They have used 500 images of each characters and experiment on 22000 character samples. They have got 78.05% accuracy on large dataset using Naive Bayes classifier.

J. Ayush Purohit and Shardul Singh Chauhan [10] mentioned six different phases for handwritten recognition are pre processing, segmentation, feature extraction, classification and post processing. After reviewing different research paper they have compared different methods for character recognition and analysed that at the stage of feature extraction and classification we should have proper method for good accuracy rate for character recognition. So there is scope for building the algorithm and increase the rate for character recognition.

K. Vaidehi Patel and Prof. Abhinay Pandya [11] have used self generated dataset from more than 100 writers with different age group. They have compared different feature extraction techniques like zoning, moments invariants, freeman chain code, Fourier transform and projection histogram. They have also compared various classification approaches like template matching, neural networks, support vector machine, k-nn classification and hidden markov model. They mentioned that Hidden markov model does not required training so, after surveying different paper and comparison with all technique they use freeman chain code for feature extraction and hidden markov method for classification.

L. Apurva A. Desai [13] considered 40 handwritten characters which are gather from 199 writers so total 7960 alphabets as a dataset for testing. In this paper Support Vector Machine (SVM) method and k-NN is used for classification. Both Method is compared for classification stage and as a conclusion SVM with polynomial kernel gives good accuracy compare to other. He has achieved average 86.66% accuracy for identification of alphabets and mentioned that identification of ૐ, ૫, ૭, ૬, ૫, and ૧ character as confusing alphabet.

M. Shailesh Chaudhari and Dr. Ravi Gulati [14] focus on segmentation of handwritten character in Guajarati script. They mentioned that selection of good segmentation technique can improve accuracy of recognition. In segmentation process due to conjunct consonants character recognition may become very difficult especially when handwriting of different people. In this paper they have divided segmentation problem into three categories like problem in line segmentation, problem in Word segmentation and problem in character segmentation. They have concluded that to solve this type of problem better algorithm can be develop and they are currently working on it.

N. Arpit Jain, Harshal Arolkar and Chirag Davda [15] proposed pattern recognition system which used for recognizing Gujarati handwritten character and they have experimented it on five characters with data set of 6750 handwritten Gujarati character. They have proposed architecture of Gujarati Handwritten Character Recognition System (GHCRS). Using this system they have achieved accuracy for dis-joint characters are 100%, 46.66%, 97.33%, 100% and 82.66% on characters Ga, Ana, La, Sha and Ha respectively. They have used multiple algorithms in proposed system for Image Segmentation, Pattern Matching and Pattern Generation.

Table 1: A Comparative study of Handwritten Gujarati Character Recognition

Paper Title	Year	Work On	Accuracy/Conclusion
A Survey of Gujarati Handwritten Character Recognition Techniques [1]	2018	Handwritten Gujarati Character	Individual characters, numbers or sample are not providing accuracy of recognizing characters at satisfactory level
Offline Handwritten Gujarati Numeral Recognition Using MLP Classifier [2]	2018	Handwritten Gujarati Numeral	95.12% on Training, 90% on Testing and 93.83% on Cross Validation
Online Handwritten Gujarati Numeral Recognition Using Support Vector Machine [3]	2018	Online Handwritten Gujarati Numeral	95% using SVM polynomial
A Review of Handwritten Character Recognition [4]	2017	Gujarati Handwritten Character Recognition	Modifiers and Conjunct Characters are complex to identify
Offline Handwritten Character Recognition for Gujarati Language [5]	2017	Offline Gujarati Handwritten Character	97.92% for SVM and 97.27% for K-NN
Survey on Offline Character Recognition for Handwritten Gujarati Text [6]	2017	Offline Gujarati Handwritten Character Recognition	Modifiers and Conjunct characters are complex to identify
Offline Handwritten Numeral Recognition using Combinational Feature Extraction Approach [7]	2017	Offline Handwritten Numeral Recognition	96.8%
A Study to Recognize Printed Gujarati Characters Using Tesseract OCR [8]	2017	Printed Gujarati Characters	More than 80%
Handwritten Gujarati Character Recognition Based On Discrete Cosine Transform [9]	2016	Handwritten Gujarati Character	78.05%
A Literature Survey on Handwritten Character Recognition [10]	2016	Handwritten Character Recognition	Feature extraction and Classification stage should have proper method for good accuracy rate for character recognition

A Survey on Gujarati Handwritten OCR using Morphological Analysis [11]	2016	Offline Gujarati Handwritten Character	Compared different methods and algorithm for OCR
Support vector machine for identification of handwritten Gujarati alphabets using hybrid feature space [13]	2015	Handwritten Gujarati Character (થ, પ, ઘ, ધ, ય, and વ)	86.66%
Segmentation Problems in Handwritten Gujarati Text [14]	2014	Handwritten Gujarati Character	Segmentation process due to conjunct consonants character recognition may become very difficult and better segmentation related algorithm can be develop
Recognition of Offline Gujarati Handwritten Disjoint Consonants using Pattern Matching [15]	2019	Disjoint Gujarati Handwritten Characters (ગ, ણ, ત, શ and ઙ)	100%, 46.66%, 97.33%, 100% and 82.66%

IV. CONCLUSION

This paper presents offline Handwritten Character Recognition and online Handwritten Character Recognition. In Gujarati language developing accurate handwritten character recognition becomes a huge challenge. There are various techniques have been proposed towards different Indian languages in offline handwritten character recognition but comparatively very less work has been found for Gujarati language. The main reason for behind it is different writing style of people. This paper defined and categorized different forms of joint characters or conjunct character. The joint characters or conjunct characters are complex to identify. Very less word has been done for joint characters which need lot of research and efficient algorithm for better accuracy.

REFERENCES

[1] Arpit A. Jain and Harshal A. Arolkar, “A Survey of Gujarati Handwritten Character Recognition Techniques”, International Journal for Research in Applied Science & Engineering Technology (IJRASET), ISSN: 2321-9653, Volume 6, Issue IX, Sep 2018

[2] Dr. Dinesh Satange, Dr. P E Ajmire and Fozia I Khandwani, "Offline Handwritten Gujarati Numeral Recognition Using MLP Classifier", International Journal of Novel Research and Development(IJNRD), ISSN: 2456-4184, Volume 3, Issue 8, Aug 2018

[3] V. A. Naik and A. A. Desai , "Online Handwritten Gujarati Numeral Recognition Using Support Vector Machine", International Journal of Computer Sciences and Engineering, ISSN: 2347-2693, Volume 6, Issue-9, Sep 2018

[4] Nikita Mehta and Jyotika Doshi, "A Review of Handwritten Character Recognition", International Journal of Computer Applications, Volume 165 – No.4, May 2017

[5] Sumit G. Trivedi and Arun Nandurbarkar, "Offline Handwritten Character Recognition for Gujarati Language", International Journal for Scientific Research and Development, ISSN: 2321-0613, May 2017

[6] Bhumika B. Patel and Hinaxi M. Patel, "Survey on Offline Character Recognition for Handwritten Gujarati Text", International Journal of Computer Applications (0975 – 8887), Volume 177 – No.6, Nov 2017

[7] Nishu Rastogi, Maitreyee Dutta and Ajay Indian, "Offline Handwritten Numeral Recognition using Combinational Feature Extraction Approach", International Journal of Innovative Research in Computer and Communication Engineering, ISSN(Online): 2320-9801, ISSN (Print): 2320-9798, Volume 5, Issue 4, Apr 2017

[8] Milind Kumar Audichya and Jatinderkumar R. Saini, "A Study to Recognize Printed Gujarati Characters Using Tesseract OCR", International Journal for Research in Applied Science & Engineering Technology (IJRASET), ISSN: 2321-9653; Volume 5, Issue IX, Sep 2017

[9] MADHUREE ARDESHANA, ANKIT K. SHARMA, DIPAK M. ADHYARU and TANISH H. ZAVERI, "Handwritten Gujarati Character Recognition Based On Discrete Cosine Transform", Proceedings of IRF-ieeeforum International Conference, ISBN: 978-93-85973-95-6, Apr 2016

[10] Ayush Purohit and Shardul Singh Chauhan, "A Literature Survey on Handwritten Character Recognition", International Journal of Computer Science and Information Technologies (IJCSIT), ISSN:0975-9646, Volume 7, 2016

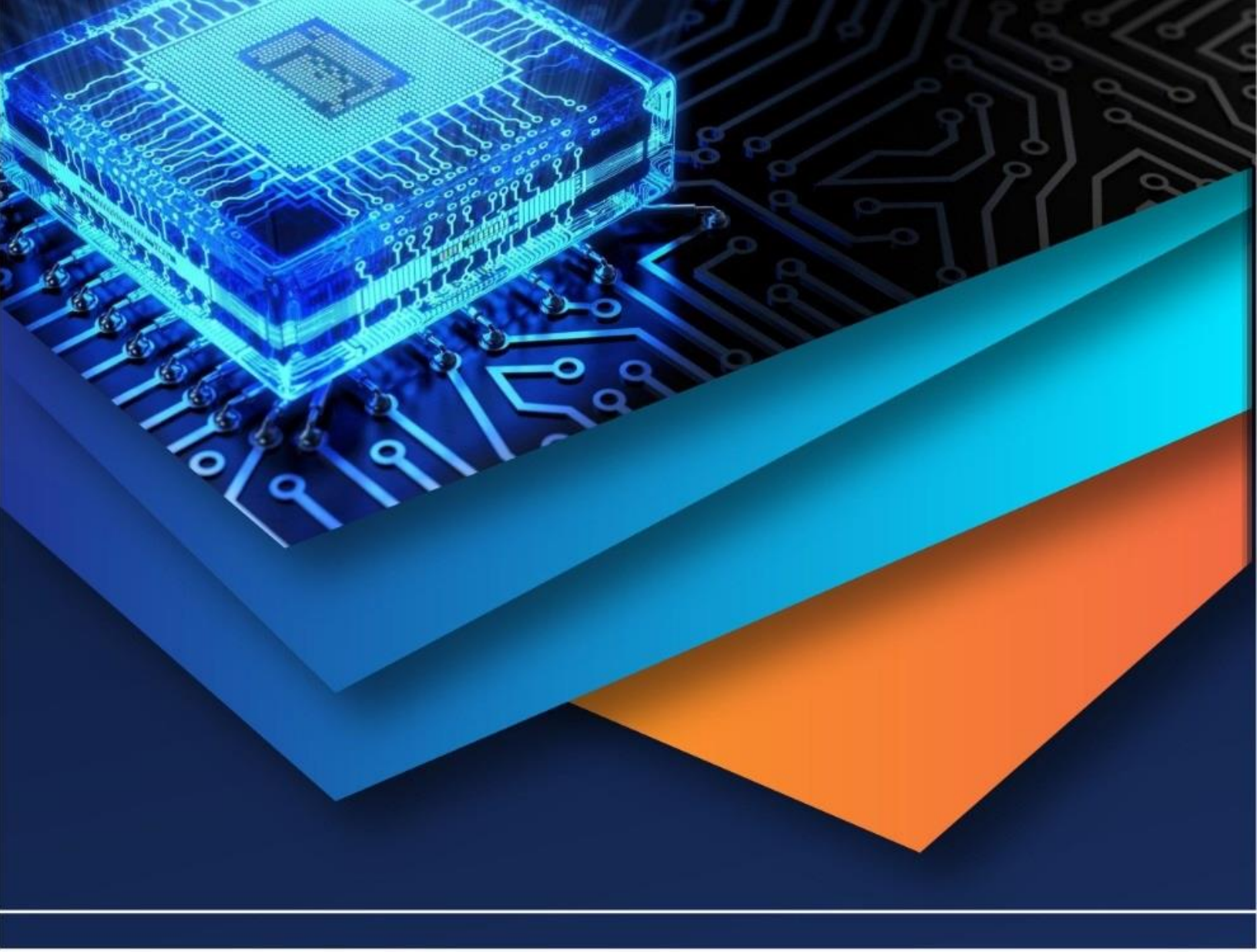
[11] Vaidehi Patel and Prof. Abhinay Pandya, "A Survey on Gujarati Handwritten OCR using Morphological Analysis", IJSRSET, print ISSN : 2395-1990, online ISSN : 2394-4099, Volume 2, Issue 2, 2016

[12] https://dol.gujarat.gov.in/writereaddata/images/pdf/guj-guj_shabdkosh_cd.pdf (For Gujarati Dictionary)

[13] Apurva A. Desai, "Support vector machine for identification of handwritten Gujarati alphabets using hybrid feature space", Springer, CSI Publications, 2015

[14] Shailesh Chaudhari and Dr. Ravi Gulati, "Segmentation Problems in Handwritten Gujarati Text", International Journal of Engineering Research & Technology (IJERT),ISSN: 2278-0181, Volume 3, Issue 1, Jan 2014

[15] Arpit A Jain , Harshal A Arolkar, Chirag S Davda, "Recognition of Offline Gujarati Handwritten Disjoint Consonants using Pattern Matching", International Journal of Recent Technology and Engineering (IJRTE), ISSN: 2277-3878, Volume 8 Issue 3, 2019



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY