International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2020): 7.803

Comparative Analysis on Handwritten Gujarati Numerals Recognition

Dr. Bharat C. Patel

Smt. Tanuben & Dr. Manubhai Trivedi College of Information Science, Veer Narmad South Gujarat University, Surat, Gujarat, India.
patelbharat99[at]yahoo.co.in

Abstract: ***The world is moving very rapidly towards digital era and therefore, any historical or government documents are also required to convert into digital form. The process of translating documents that are in the form of paper into digitized form is called optical character recognition (OCR) which is the field of research in pattern recognition. Character or numeral recognition is a most important field of pattern recognition. In the current scenario, Gujarati numeral OCR play significant role in various applications like reading numerals written on bank cheque which is either printed or handwritten form, recognition of numeric entries filled in the form or even other related applications. This paper deals with research work done in printed and handwritten numeral recognition on Gujarati script. Obviously, the problem of handwritten numerals has variety of challenges like numerals are not of the same size, style, orientation and thickness relative to printed numerals. Therefore, they are complicated to recognize rather than typed Gujarati numerals. Compare to other regional Indian languages, the research work completed in Gujarati language is not up to the mark. This paper focuses on literature study on Gujarati numeral recognition that helps learners to provide basic structure of recognition system and also guide them how to proceed further in recognition task.***

Keywords: Online Gujarati numeral recognition, offline Gujarati numeral recognition, feature extraction method, classifiers

1. Introduction

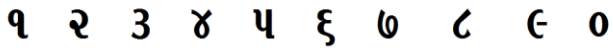
The process of recognizing either typed or handwritten text and translate it into an editable digitized form is known as character recognition which is a field of natural language processing. Researchers are continuously attempting to develop a user-friendly interface that helps them to recognize both printed and handwritten characters. As per the recognition scheme, the character recognition system can be categorized into two classes:

- Online character recognition and
- Offline character recognition.

In online character system, either characters or numerals are entered on active supervision of a user to a computer or handheld device by means of a pen or stylus or touch screen device. In the later system, handwritten documents in the form of paper are scanned and further analyze by computer to identify the digitized characters or numerals. Literature study shows that notable work is done in handwritten offline Gujarati numeral recognition but very little work is done in online Gujarati numeral recognition [1].

Still, no any universal algorithm is developed or even possible for all Indian language, because each language has its own characteristics and limitations. In Gujarati script, handwritten character recognition is one of the most important and complex tasks due to variety in writing styles and curve of large and complex character set. Gujarati is an Indo-Aryan type of language and officially most spoken language of Gujarat state. Typed and handwritten Gujarati numeral set includes numerals as shown in figure 1.

(a)



(b)

Figure 1: Sample image of Gujarati numerals
(a) Handwritten Gujarati numerals (b) Typed Gujarati Numerals]

This paper is prepared as follows: Section 2 explains the literature review regarding automatic recognition of numerals written in Gujarati language while Section 3 compares the various handwritten digit recognition techniques. Section 4 emphasizes some of the uncover challenges for Gujarati handwritten recognition. Finally, section 5 concludes conclusions and future work.

2. Early attempt in Gujarati language OCR

In the literature, little bit of work is done for online/offline typed or handwritten numerals recognition in Gujarati scripts. Some of the researchers have been attempted to work on either online or offline numerals recognition in Gujarati script and obtained good success rate by the proposed algorithm is reported.

Niak and Desai [1] proposed an online handwritten Gujarati word recognition system. They considered 101 unique classes of Gujarati language for proposed system and utilize novel features as normalized and zoning-based chain code features. They achieved an average accuracy of 95.3% for individual characters, 91.5% for individual words, and 83.3% for sentences.

Desai [2] developed a system for recognition of offline handwritten numerals in Gujarati script. The proposed system utilize various techniques in pre-process phase and a multi layered feed forward neural network is applied to classify the numerals. The author obtained correct success

rate of 81.66% for identification of Gujarati handwritten numerals.

Maloo and Kale [3] proposed a system that uses support vector machine classifier to recognize handwritten Gujarati numerals. They created four feature sets and obtained the recognition rate of 91% approximately.

Niak and Desai [4] proposed an algorithm which recognizes online handwritten Gujarati numerals. The feature set of 22 different structural and statistical features are used in the proposed system. Support Vector Machine is used as a classifier with a linear, polynomial, and RBF kernel and achieved an accuracy of 92.60%, 95%, and 93.80% respectively.

Goswami and Mitra [5] had developed a database for offline handwritten numeral in Gujarati language. They collected 14, 000 samples of handwritten Gujarati characters from 140 peoples in order to create database. The persons having dissimilar education stream, age group and work culture were considered for the purpose samples data collection. Low-level stroke (LLS) features such as line segments, curve segments, endpoints and junction points are extracted using template matching approach. Initially, they performed classification using k-nearest neighbour (k-NN) after that statistically advance SVM with radial basis function (RBF) kernel is used to improve the results. The average test accuracy 98.46% and 98.65% were received for Gujarati and Devanagari database respectively.

Patel and Kaystha [6] proposed a system that recognized offline handwritten Gujarati numerals. The authors utilize various features of numerals like straight-line, number of end edges present in zones, hole, etc to recognize numerals in Gujarati script and achieved good accuracy rate using proposed system.

Niak and Desai [7] suggested two-layer classification techniques i. e. in layer one SVM with RBF kernel and in layer two k-NN classifier is used to identify online handwritten characters in Gujarati script. First and second order derivative of a hybrid features of pixel values, zoning, and normalized chain code are delivered to two-layer classifier. The proposed system returns response in 0.095 seconds per stroke and received an average accuracy rate 94.65% to recognize handwritten character.

In [8], Patel and Desai proposed the system which segment printed Gujarati numerals present in an image and recognized them using k-NN classifier. In addition, they declared that no any work is reported to segment and recognize Gujarati numerals from the image. The experiment performed on different type of images having numerals in it and obtained a success rate about 95.80%. Moreover, they also argued that the proposed model worked well on the images having attributes like coloured, textured, map and even on magazine cover pages and noisy image.

Pareek, J. et al. [19] proposed handwritten character recognition system for Gujarati script. The proposed system employed a supervised classifier as Convolutional Neural Networks (CNN) and Multi-Layer Perceptron (MLP) and achieved recognition rate 97.21% and 64.48% respectively.

In [21], Gupta, D. and Soumen Bag proposed multilingual handwritten numerals recognition system. The system is script independent which has a capability to recognize handwritten numerals written in eight different scripts. The proposed system received correct accuracy rate of 96.23% collectively for all the eight scripts, namely English, Hindi, Bangla, Odia, Telugu, Gujarati, Punjabi and Arabic.

P E Ajmire and Fouzia I Khandwani [24] proposed a system that recognizes handwritten Gujarati numerals. They selected the features of numerals as convex area, filled area, Euler number and Eccentricity. The methods based on artificial neural network (ANN), support vector machine (SVM) and naive Bayes (NB) classifier are applied on handwritten Gujarati numerals recognition and obtained average accuracy of recognition is 98.33%.

3. Comparative Analysis

As per the literature review, it was seen that the offline handwriting is relatively more challenging than typed/printed characters. Moreover, Gujarati numerals have sharp curve and the hand writing style of numerals is differ from person to person and therefore recognition of handwritten Gujarati numerals is more complicated and relatively challenging task. Researchers have find out various features of numerals and attempted to recognize them. Table1 illustrates comparative analysis of Gujarati numeral recognition that includes the features and classifier used by researchers and success rate obtained in their proposed work.

Table 1: Comparative analysis of Gujarati numeral recognition

Researchers	Features used	Classifier used	Recognition mode	Success rate
Niak and Desai [1]	a percentage of active pixels in different zones, normalized chain code, and zoning-based chain code	SVM with RBF	Online handwritten	Characters-95.3%, Words-91.5% and Sentences-83.3%
Gohel et al [9]	Low level stroke feature and directional feature	K-NN	Online handwritten	Numerals – 95%, Characters – 93% Mixed – 90%
Prasad and Kulkarni [10]	Gabor Phase XNOR Pattern, Pattern descriptor, Contour Direction Probability Distribution Function, Autocorrelation	Adaptive Neuro Fuzzy Classifier (ANFC)	Offline handwritten character	68%

Macwan and Vyas [11]	Freeman chain code, Hu's (4 order), center of mass	SVM	Offline handwritten character	87.29%
Vyas and Goswami [12]	Discrete Cosine Transform coefficients	K-NN	Offline handwritten digits	93.60%
Mendapara and Goswami [13]	directional features	k-NN	Offline handwritten digits	88%
Patel and Desai [8]	Aspect ratio and Extent	template matching	Offline printed numerals	95%
Baheti M. J. Kale K. V. [22]	affine invariant moments	SVM, Gaussian distribution function, K-NN	Offline handwritten numerals	92.28%
Mikita Gandhi et al. [23]	wavelet decomposition	KNN	offline Handwritten Numeral	98.30%

Notes: Some of the researchers have used multiple classifiers but the classifier with best recognition accuracy has been mentioned.

4. Future Scope

The work reported on Gujarati handwritten numeral recognition can be extended in several directions if the following issues are taken into account. Few of them are listed below:

(i) Improved pre-processing

For Gujarati handwritten numeral recognition task, the pre-processing stage essentially involved in some pre-processing sub task like binarization, foreground and background noise elimination, size normalization, skew and slant corrections, etc. Some handwritten numeral recognition methods also need that the numeral images to be segmented and thinned [8] (Patel and Desai, 2014) that will finally helpful to improve feature extraction.

(ii) Extraction of competent features

Feature extraction usually provides the extraction of the useful information from the input numeral image, which in turn make possible for better classification of different numeral images in the feature space. It can be observed from the literature study that there are some general features like geometry features (such as vertical line, horizontal line, left diagonal and right diagonal lines, area and Euler number), Character Profile (such as vertical, horizontal, left and right diagonal profile of objects), structural and statistical features etc. have been repeatedly used by the researchers for Gujarati numeral recognition. Most of the researchers have been used general features mentioned earlier in this paragraph but few of them had applied features like zoning and chain code directional features (Niak et al [4]), modified chain code (Archana et al. [12]), intersection, shadow feature, chain code histogram and straight line fitting features (Arora et al [16]) etc which in turn provided better classification results than the former general features. Sharma, A. K. et al. [20] have described the problem of handwritten character recognition in Gujarati script. They proposed a system with three new features to represent handwritten characters in Gujarati language. The features are extracted based on normalized cross correlation, zone pattern matching and structural decomposition and received highest accuracy rate of 98.77% is obtained while using structural decomposition based features.

(iii) Fusion of features

To get improve result of character recognition system, the researchers are trying to hybrid more than one feature so that it can perform universally any situation. But at present there is no any universal method has been found out. Some researchers have tried to fuse features and applied them in their proposed work. Sharma et al [15] combine Zone-based, projection profiles-based and chain code-based features and are employed as individual feature. Ajay Indian and Karamjit Bhatia [17] proposed a feature extraction algorithm based on combinational approach which mixes up the feature vectors for instance Zernike complex moment features, Gradient features and Wave based features to improved recognition performance of the system. They utilized samples of numerals in Hindi script from 0 to 9 in order to train and test the system. Different feature vectors of Zernike complex moments (ZCM), directional gradient histogram (DGH) and Wave features (WF) are supplied to the Back-propagation based Neural Network classifiers for training and get correct recognition rate of 79.7%, 92.7% and 73% are obtained respectively. In addition, integration of feature vectors of ZCM, DGH, and WF, they achieved a higher recognition rate of 96.4% for isolated Hindi Numerals. Danveer Rajpal et al [18] proposed a system which adopted fused features for Hindi character recognition. The DCNN-based features are compound with the features, i. e. handcrafted features, obtained from Bi-orthogonal discrete wavelet transform. The size of feature vector is reduced by the Principal Component Analysis method. In addition, the fusion features are observed with very familiar classifiers that are Multi-Layer Perceptron (MLP) and Support Vector Machine (SVM) to reduce the recognition cost by 84.37%. The proposed system obtained considerable scores of recall, precision and F1-measure 98.67%, 98.78% and 98.69% and overall average correct rate of 98.73% is achieved for recognition of character.

(iv) Fusion of different classifiers

The purpose of upcoming research should be to find the best combination of classifiers for intend of numeral recognition. In the circumstances, where no any single classifier can provide a satisfactory level of classification accuracy, it becomes requisite to get better results of different classifiers in order to accomplish the best accuracy. Following selection of complementary features, combining multiple classifiers may also significantly improve the reliability of the Gujarati handwriting numeral recognition system. Some of the works described in [14] Patel and Desai combined the results of their classification results to accomplish better recognition accuracies but a general combination scheme designed for Gujarati handwriting numeral recognition system will be of great importance. Moreover, it can be observed in [16] that mixture of individual classifier outputs conquers lacks of features and trainability of single

classifier. Outputs from more than one classifiers can be come together to generate a more precise result. Arora et al [16] were employed weighted majority voting technique for merging the classification decision which is obtained from four Multi Layer Perceptron (MLP) based classifier and received the correct overall recognition rate is 92.80%.

(v) Necessity of Benchmark databases

Creation of standard handwritten numeral recognition test database is crucial for Gujarati as well as other Indic scripts. The research works done earlier on recognition of Gujarati handwritten numerals were described on the basis of databases composed in the laboratory. In order to assemble a pragmatic system, researchers need real-life digit samples from different writers belonging to miscellaneous backgrounds, educational levels, gender and age groups. There are no standard databases that are freely available for Gujarati handwriting numeral recognition.

5. Conclusion

The literature study includes work related to the recognition of handwritten Gujarati numerals. The comparisons of all the significant proposed work have been presented in tabular form. This comparison is done based on the feature set, classifier, recognition mode and the reported accuracy rates. The following issues for Gujarati handwriting numeral recognition system are noticed:

a) Most of the research work has been done on smaller subset of dataset created by them.
b) Time and space complexity related problems has not considered in most of the research works. Few of the researchers such as Niak and Desai [1] have considered the average processing time for individual characters is 0.071 seconds.
c) Some feature extraction methods are yet restricted to capture either local or global information and so, the lack of combination of both the scopes will be vital for handwritten numeral recognition system written in any official Indic scripts.

This analysis point out that how the research tendency has been changed over the years, study the various techniques being utilized for classification, and draw attention on to the limitations of the existing Gujarati handwriting numeral recognition systems.

References

[1] Vishal A. Naik and Apurva A. Desai, ‘Online Handwritten Gujarati Word Recognition’, International Journal of Computer Vision and Image Processing, Volume 9, Issue 1, January-March 2019, pp.35-50.
[2] Apurva A. Desai, ‘Gujarati handwritten numeral optical character reorganization through neural network’, Pattern Recognition 43 (2010) 2582–2589.
[3] Mamta Maloo and K. V. Kale, ‘Support Vector Machine based Gujarati Numeral Recognition’, International Journal on Computer Science and Engineering (IJCSE), Vol.3 No.7 July 2011, pp.2595-2600.
[4] V. A. Naik and A. A. Desai, “Online Handwritten Gujarati Numeral Recognition Using Support Vector Machine”, International Journal of Computer Sciences and Engineering, Vol.-6, Issue-9, Sept.2018, pp.416-421.
[5] Goswami M. M. and Mitra S. K. (2015) ‘Offline handwritten Gujarati numeral recognition using low-level strokes’, Int. J. Applied Pattern Recognition, Vol.2, No.4, pp.353–379.
[6] Patel Bharat C. and Kaysth Manish M., “Recognition of Offline Handwritten Gujarati Numerals”, i-manager’s Journal on Information Technology, Vol.5 l No.5, December 2016-February 2017, pp.14-18.
[7] Vishal A. Naik and Apurva A. Desai, “Online Handwritten Gujarati Character Recognition: Two-Layer Classification Approach”, International Journal of Research in Advent Technology, Vol.7, No.6, June 2019, pp.184-190.
[8] Jagin M. Patel and Apurva A. Desai, “Segmentation and Recognition of Gujarati Printed Numerals from Image”, International Journal of Engineering Research & Technology (IJERT), Vol.3 Issue 2, February – 2014, pp.1012-1020.
[9] C. C. Gohel, M. M. Goswami and V. K. Prajapati, "On-line handwritten Gujarati character Recognition using low level stroke", 2015 Third International Conference on Image Information Processing (ICIIP), 2015, pp.130-134, doi: 10.1109/ICIIP.2015.7414753.
[10] J. R. Prasad and U. V. Kulkarni, "Gujrati Character Recognition Using Adaptive Neuro Fuzzy Classifier", 2014 International Conference on Electronic Systems, Signal Processing and Computing Technologies, 2014, pp.402-407, doi: 10.1109/ICESC.2014.79.
[11] S. J. Macwan and A. N. Vyas, "Classification of offline Guajarati handwritten characters, " 2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2015, pp.1535-1541, doi: 10.1109/ICACCI.2015.7275831.
[12] N. Vyas and M. M. Goswami, "Classification of handwritten Gujarati numerals, " 2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2015, pp.1231-1237, doi: 10.1109/ICACCI.2015.7275781.
[13] M. B. Mendapara and M. M. Goswami, "Stroke identification in Gujarati text using directional feature", 2014 International Conference on Green Computing Communication and Electrical Engineering *(ICGCCEE)*, 2014, pp.1-5, doi: 10.1109/ICGCCEE.2014.6922264.
[14] Chhaya Patel and Apurva Desai, “Gujarati handwritten character recognition using hybrid method based of binary tree classifier and K-nearest neighbor’, International Journal of Engineering Research & Technology (IJERT), Vol.2 Issue 6, June – 2013, pp.2337-2345.
[15] Sharma, A. K., Thakkar, P., Adhyaru, D. M. and Zaveri, T. H. (2017) ‘Gujarati handwritten numeral recognition through fusion of features and machine learning techniques’, Int. J. Computational Systems Engineering, Vol.3, Nos.1/2, pp.35–47.
[16] Arora, S., Bhattacharjee, D., Nasipuri, M., Basu, D. K., & Kundu, M. (2008), “Combining Multiple Feature Extraction Techniques for Handwritten

Devnagari Character Recognition", 2008 IEEE Region 10 and the Third International Conference on Industrial and Information Systems. doi: 10.1109/iciinfs.2008.47984.

[17] Ajay Indian and Karamjit Bhatia, "A Combinational approach of Feature Extraction for Offline Handwritten Hindi Numeral Recognition", International Journal of Computer Sciences and Engineering, Vol.-6, Issue-5, May 2018, pp.1-8.

[18] Rajpal, D.; Garg, A. R.; Mahela, O. P.; Alhelou, H. H.; Siano, P. A Fusion-Based Hybrid-Feature Approach for Recognition of Unconstrained Offline Handwritten Hindi Characters. *Future Internet* 2021, *13*, 239. https: //doi. org/10.3390/fi13090239.

[19] Pareek, J., Singhania, D., Kumari, R. R., & Purohit, S. (2020). Gujarati Handwritten Character Recognition from Text Images. Procedia Computer Science, 171, pp.514–523. doi: 10.1016/j. procs.2020.04.05.

[20] Sharma, A. K., Thakkar, P., Adhyaru, D. M. and Zaveri Tanish H., Handwritten Gujarati Character Recognition Using Structural Decomposition Technique. Pattern Recognit. Image Anal.29, 325–338 (2019). https: //doi. org/10.1134/S1054661819010061.

[21] Gupta, D., & Bag, S. (2020). CNN-based multilingual handwritten numeral recognition: A fusion-free approach. Expert Systems with Applications, 113784. doi: 10.1016/j. eswa.2020.11378.

[22] Baheti M. J. and Kale K. V., "Gujarati Numeral Recognition: Affine Invariant Moments Approach", 1st International Conference on Recent Trends in Engineering & Technology, Special Issue of International Journal of electronics, Communication & Soft Computing Science & Engineering, Mar-2012, pp.140-146.

[23] Mikita Gandhi, Dr. V. K. Thakar, Dr. H. N. Patel, "Handwritten Gujarati Numeral Recognition using Wavelet Transform", Journal of Applied Science and Computations, Volume VI, Issue IV, April 2019, pp.2699-2707.

[24] P E Ajmire and Fouzia I Khandwani, "Feature extraction and recognition of handwritten Gujarati numerals", Journal of Emerging Technologies and Innovative Research, January 2019, Volume 6, Issue 1, pp.617-621.

Author Profile

Dr. Bharat C. Patel is an Associate Professor and I/C Principal in Smt. Tanuben and Dr. Manubhai Trivedi College of Information Science at Surat, Gujarat, India. He has more than 20 years of teaching experience. He has received B. Sc. in Physics from the Veer Narmad South Gujarat University in 1995. He received his MCA, M. Phil., and Ph. D. in Computer Science from the same University in 1999, 2013, and 2020 respectively. He is a member of the BOS of computer Science and information technology. He is a reviewer in a reputed journal. He has author four books in the field of computer science. He is a member of the Ph. D. Doctoral Progress Committee (Computer Science) at Bhagwan Mahavir University. His areas of current research interest are speech processing and recognition, image processing, machine learning, and artificial intelligence.