1st International Conference on Recent Trends in Engineering & Technology, Mar-2012
Special Issue of International Journal of electronics, Communication & Soft Computing Science & Engineering, ISSN: 2277-9477

Gujarati Numeral Recognition: Affine Invariant Moments Approach

Baheti M. J. **Kale K. V.**

***Abstract* — Handwriting recognition is transformation of text from spatial form of graphical marks to the similar symbolic representation. Automated recognition of text has been an active subject of research since the early days of computers. Gujarati, an Indo-Aryan language is spoken by about 50 million people in the Indian states and across the globe. The differences between the scripts are primarily in their written forms where different combinations come into play. In this paper an attempt is made to recognize the offline isolated handwritten Gujarati numerals. Affine invariant moments is used extracting the features. The extracted feature set is treated with Principal Component Analysis (PCA), Support Vector Machine (SVM), K-Nearest Neighbor (KNN) and Gaussian distribution function to classify the numerals. The comparison of all the classifiers is made it has been seen that SVM classifier has shown better results as compared to PCA, KNN and Gaussian distribution function classifier.**

***Key Words* — Affine Invariant Moments, Gaussian Distribution function, Gujarati, K-Nearest Neighbor, Principal Component Analysis, Support Vector Machine.**

I. INTRODUCTION

Handwriting recognition can be defined as the task of transforming text represented in the spatial form of graphical marks into its symbolic representation [1]. Since the early days of computers, automated recognition of text has been an active subject of research. This is useful for making digital copies of handwritten documents, and also in many automated processing tasks, such as automatic mail sorting or cheque processing. In automated mail sorting, letters are directed to the correct location by recognition of the handwritten address. Within the background of script recognition, it possibly will be connotation to study the distinctiveness of a variety of writing systems and the structural properties of the characters used in most significant scripts of the globe. There is only some degree of achievement in handwriting recognition, virtually for isolated and neatly hand-printed characters and word for limited vocabulary [2].

Approximately twenty-two official languages are used across various regions of India[3-6]. Each language has the variety representing its peculiarity as well as shares some of the similarities with other languages. Gujarati language is written by using Gujarati script which in turn is derived from Devanagari. Fig. 1 shows the numerals belonging to Gujarati language.

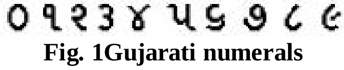


Fig. 1Gujarati numerals

Gujarati numerals show some same appearances like other numerals in Devanagari. Numerals like 0, 2, 3, 4, 7 and 8 are same as that in Devanagari but numeral 1 has. Researchers from Maharaja Sayajirao University, Baroda, India had proposed Gujarati text editor, along with spell check feature. They had used Discrete Cosine Transform as feature extractor and they worked in line to achieve almost 97% recognition rate [7]. Antani [8], Desai [9], Dholakia [10, 11] reported studies for Gujarati script and have achieved 67%, 82% and 95-97% of accuracy respectively for printed text and handwritten text. Among these only Desai [9] has given the approach for handwritten numerals in Gujarati script.

In this paper we have made an attempt to recognize the offline isolated handwritten Gujarati numerals. For feature extraction we have used affine invariant moments. After extracting features, the feature set is then treated with PCA, SVM, KNN and Gaussian distribution function to classify the numerals. After making the comparison of all the classifiers, it has been seen that SVM classifier has shown better results as compared to PCA, KNN and Gaussian distribution function classifier.

This paper is organized in following sections; Section 2 describes brief literature survey done for Indian languages recognition. Section 3 details the various steps taken for database preparation. Section 4 describes preprocessing and algorithm which we have used to implement the paper. Section 5 elaborates the feature extraction done. Section 6 describes Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Gaussian distribution function and PCA (Principal Component Analysis). Section 7 elaborates the experimental results while section 8 details the conclusion of work done.

II. LITERATURE REVIEW

A. Indian Scripts

Veena presented a survey of different structural techniques used for feature extraction in OCR of different scripts [12]. Connel [13] reported an online Devanagari script recognition attempted with 86.5% accuracy on a database of 20 writers. Sinha & Bansal [14] achieved 93% performance on individual characters. Printed Devanagari character recognition has been attempted based on K-nearest neighbor (KNN) and Neural Networks classifiers [15]. Pal & Chaudhuri have attempted OCR for two scripts, Bangla and Devanagari in [16]. Pal & Chaudhuri [17] reported a complete OCR system for printed Devanagari. A structural feature-based tree classifier recognizes modified and basic characters, while compound characters were recognized by hybrid approach combined with structural and run based template features. The method reported about 96% accuracy.

In the reported work [18] an effective method incorporated edge directions histograms and splines along with PCA were used for recognition of isolated handwritten

Devanagari numerals. A comparative study of Devanagari handwritten character recognition using twelve different classifiers and four sets of feature was presented by Pal [19]. In the reported work [20] several approaches that deal with problem of recognition of numerals/character were compared using SVM and KNN on handwritten as well as on printed character.

A zone-based hybrid feature extraction algorithm scheme for recognition of off-line handwritten numerals of Kannada and Tamil scripts. The nearest neighbor classifier obtained 97.75% and 93.9% of recognition rate for Kannada and Tamil numerals respectively. They obtained 98.2% and 94.9% of recognition rate for Kannada and Tamil numerals respectively using support vector machine classifier [21]. In the proposed work [22], the author found 4 directional images using Gabor wavelets from the dynamically preprocessed original images. Then extracted moment features from them. The comparison of moment features of 4 directional images with original images when tested on Multi Layer Perceptrons with Back Propagation Neural Network showed an average improvement of 13% from 72% to 85%. The mean performance of the system with these two features together was 92%.

In the reported work [23], a system to recognize isolated handwritten characters in Gurmukhi characters has been developed. The feature extraction method, Zoning and two classification methods, k-nearest neighbor, SVM (support vector machines) had been used procuring results as 72.54%, 72.68%, 73.02% and 72.83% respectively.

Hybrid scheme was proposed for Gurmukhi script by Kaur [24]. The concept of water reservoir as well as Feature Extraction was used. Singh [25] have used some statistical features like zonal density, projection histograms (horizontal, vertical and both diagonal), distance profiles (from left, right, top and bottom sides). In addition, they have used background directional distribution (BDD) features. The highest accuracy obtained was 95.04% as 5-fold cross validation of whole database using zonal density and background distribution features in combination with SVM classifier used with RBF kernel among SVM, K-NN and PNN classifiers.

Chaudhari [28] used preprocessing techniques like skew correction, line segmentation, zone detection, word and character segmentation and then the combination of stroke and run-number based and water reservoir based features were used as classifiers. They achieved 96.3% of accuracy. The features of Oriya OCR developed at the Indian Statistical Institute, Kolkata were similar to the Bangla OCR developed by the same team Chaudhuri [26, 28]. On average, the system reported an accuracy of about 96.3%. Adaptation of Bangla OCR to Assamese was also reported in [27]. Character-level accuracy of about 95% was reported. Mohanty and Behera described a complete OCR development system for Oriya script [29]. Recognition rate of 98.40% had been reported in [30]. In the work [31], Saumya used methods for feature extraction such as dimension measurement like height and width of the image, storage details like file size and vector size. They computed centroid of the character image in horizontal and vertical coordinates, histogram of luminance for each gray level value. Lastly pixel counts in the image in total and in each row and columns separately were calculated. After all the relevant features were extracted the corresponding numeric values were stored in the database. They used Back propagation Neural Network for efficient recognition where the errors were corrected through back propagation and rectified neuron values were transmitted by feed-forward method in the neural network of multiple layers.

In the reported work [32], the algorithm consisted of unsupervised clustering method and a supervised classification technique that reported acceptable classification accuracies on both the training and test sets of the present database. Kohonen self organizing maps (KSOM) along with a fine-tuning method, that uses global features was developed for a subset of Tamil alphabet giving a recognition accuracy of 86% [33]. Bremananth & Prakash [34] trained five sets of each Tamil numeral and tested un-trained Tamil numerals are tested against the trained numerals. They pre-processed the images and extracted features in terms of rows and columns signal variations from the binarized image. In another proposed method [35], Gabor filters and support vector machines (SVM) were used for English and Tamil. The average accuracy of recognition for English was 97% and for Tamil it was 84%.

In the work [36], Tamil font recognition was done based on global texture analysis. They employed support vector machines (SVM) to identify various fonts in Tamil and feature vectors were extracted by using of Gabor filters. The SVM classifier showed an average accuracy of 92.5%. Rajashekararadhya & Ranjan [37] proposed a system on zone-based feature extraction algorithm for the recognition of numerals of Kannada, Tamil, Telugu and Malayalam scripts. 300 features were extracted for classification and recognition and obtained a recognition rate of 98.05%, for Kannada numerals, 95.1% for Tamil numerals, 97.2% for Telugu numerals and 95.7% for Malayalam numerals using support vector machine.

Jawahar [38] proposed a Bilingual OCR for Hindi-Telugu documents. It was based on Principal Component Analysis followed by support vector classification. An overall accuracy of approximately 96.7% was reported.

Lakshmi & Patvardhan [39] presented recognition of basic Telugu symbols. Feature vector was computed out of a set of seven invariant moments from the second and third order moments. Recognition was done using k-nearest neighbor algorithm on these feature vectors. In the reported work by Pratap Reddy [40] structural features of the syllable and the component model were combined to extract middle zone components. The shape of the middle zone components was closely related to a circle whereas other components were found with different topological features. Recognition rate of 99 percent was observed with the proposed method. Anuradha developed a Telugu optical character recognition system for a single font. A 2-stage classifier with first stage identified the group number of the test character, and a minimum-distance classifier at the second stage identified the character. Recognition accuracy of 93.2% was reported [41].

A modified quadratic classifier based scheme had been proposed for recognition of off-line handwritten numerals of Devnagari, Bangla, Telugu, Oriya, Kannada and Tamil scripts. The bounding box of a numeral was segmented into blocks and the directional features were computed in each of the blocks. These blocks were then down sampled by a Gaussian filter and the features obtained from the down sampled blocks were fed to a modified quadratic classifier for recognition. After using five-fold cross validation technique accuracy was 99.56%, 98.99%, 99.37%, 98.40%, 98.71% and 98.51% for Devanagari, Bangla, Telugu, Oriya, Kannada, and Tamil scripts, respectively [30].

B. Gujarati Script

Gujarati, an Indo-Aryan language is spoken by about 50 million people in the Indian states of Gujarat, Maharashtra, Rajasthan, Karnataka and Madhya Pradesh, and also in Bangladesh, Fiji, Kenya, Malawi, Mauritius, Oman, Pakistan, Reunion, Singapore, South Africa, Tanzania, Uganda, United Kingdom, USA, Zambia and Zimbabwe. Gujarati found its origin in the ancient Brahmi script, which is phonetic in nature. The alphabet in various Indian languages may vary somewhat but they all share a common phonetic structure. The differences between the scripts are primarily in their written forms where different combinations come into play [1-4, 42].

Antani & Agnihotri [8] in 1999 have given the primitive effort to Gujarati printed text. For classification the author had used two classifiers, K-NN classifier and minimum hamming distance classifier. The best recognition rate was for 1-NN for 600 dimensional binary features space i.e. 67% 1-NN in regular moment space gave 48% while minimum distance classifier had the recognition rate of 39%. In 2005, Dholakia [10] have presented an algorithm to identify various zones used for Gujarati printed text. In the algorithm they have proposed the use of horizontal and vertical profiles. They have identified these zones by slope of lines created by upper left corner of rectangle created by the boundaries of connected components from line level and not word level the 3 different document images, 20 lines were extracted where 19 were detected with correct zone boundary.

Dholakia [11] attempted to use wavelet features, GRNN classifier and KNN classifier on the printed Gujarati text of font sizes 11 to 15 with styles regular, bold and italic by finding the confusing sets of the characters producing 97.59% and 96.71% as their respective recognition rates. Kayasth & Patel [43] proposed a system for recognition of offline computer generated and printed Gujarati characters using continuous GCRHMM for different font sizes of Gujarati characters. Chaudhari [44] have described a system for recognition of offline multi-font computer generated and machine printed Gujarati numerals using correlation based template matching where a numeral is identified by analyzing its shape and comparing its features that distinguish each numeral. In the work [45-46] the author proposed a simple yet robust structural solution for performing character recognition in Gujarati. They used Template matching algorithm that comprised of Template Classification, Correlation Analysis and Computation of Cross Correlation Coefficient with average overall recognition rate of 71.66 %.

In the reported work of E.Hassan [47], multiple kernels learning (MKL) was used for Gujarati character recognition. The authors applied three different feature representations for symbols obtained after zone wise segmentation of Gujarati text. The MKL based classification was proposed, where the MKL was used for learning optimal combination of different features for classification.

Desai [9] have used four profile vectors as an abstracted feature of identification of digit. Five more patterns for each digit are created in both clockwise and anticlockwise directions with the difference of 2degrees each up to 10◦. A feed forward back propagation neural network was used for Gujarati numeral classification and a score of 81.5% recognition rate was recorded.

The extension of this work was carried out by maintaining two aspects. One was the information of subdivision of the skeletonized image and second was the aspect ratio of the image before converting it into skeleton. k-NN classifier for classification of Gujarati handwritten numerals gave accuracy of 96.99% for training set and 92.783% for unseen data [48].

III. GUJARATI DATABASE

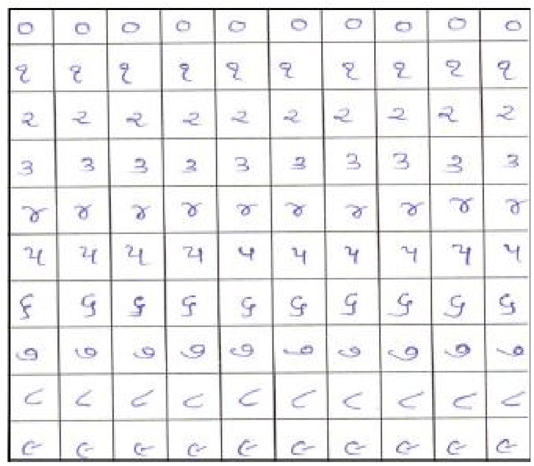


Fig. 2 Sample sheet for database collection

At present there is no standardized database available for Gujarati handwritten numerals and for the identification of Gujarati handwritten numerals the algorithm required to have database, so primarily the database has been formed. Data was collected from people of different age groups, belonging to different profession, illiterate but knowledge of writing Gujarati irrespective of gender. From such diversity of group, ten sample of each digit from 16 persons was collected on a specially designed datasheet as shown in Fig. 2 for data collection.

IV. PREPROCESSING AND ALGORITHM

The various preprocessing steps and the scheme used for implementing this paper, is shown in Fig 3. The algorithm identifies the handwritten Gujarati numerals based on iterative approach. It identifies more than one numeral. For implementing this paper, we have considered the handwritten database of sample size 1600 (800 for training and 800 for testing). The overall recognition rate if finally achieved by ratio of sum of the correctly recognized numerals by total numerals for SVM, Gaussian distribution function, KNN as well as PCA.

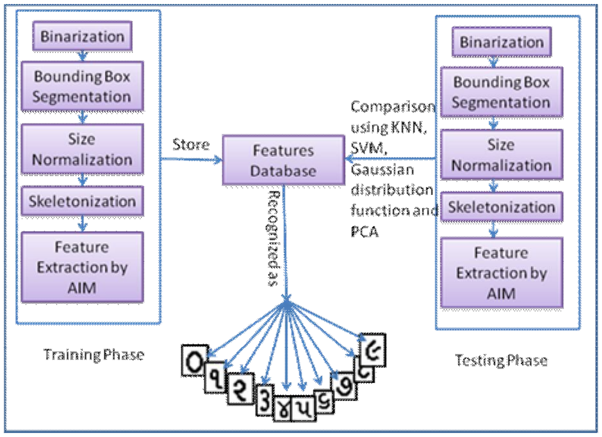


Fig. 3 Algorithm for recognition of Gujarati Numeral

V. FEATURE EXTRACTION

The affine invariant moments [49-51] are derived for each of the numeral image as follows. The AMIs is invariant under general affine transformation

$$\left.\begin{aligned} u &= a_0 + a_1 x + a_2 y \\ v &= b_0 + b_1 x + b_2 y \end{aligned}\right\} \ldots\ldots\ldots\ldots\ldots (1)$$

where, (x, y) and (u, v) are coordinates in the image plan before and after the transformation , respectively.

The basic affine invariant moments are given below:

$$\left.\begin{aligned} I_1 &= (\mu_{20}\mu_{02} - \mu_{11}^2)/\mu_{00}^4 \\ I_2 &= (\mu_{30}^2\mu_{03}^2 - 6\mu_{30}\mu_{21}\mu_{12}\mu_{03} + 4\mu_{30}\mu_{12}^3 + 4\mu_{03}\mu_{21}^3 \\ &\quad - 3\mu_{21}^2\mu_{12}^2)/\mu_{00}^{10} \\ I_3 &= (\mu_{20}(\mu_{21}\mu_{03} - \mu_{12}^2) - \mu_{11}(\mu_{30}\mu_{03} - \mu_{21}\mu_{12}) \\ &\quad + \mu_{02}(\mu_{30}\mu_{12} - \mu_{21}^2))/\mu_{00}^7 \\ I_4 &= (\mu_{20}^3\mu_{03}^2 - 6\mu_{20}^2\mu_{11}\mu_{12}\mu_{03} - 6\mu_{20}^2\mu_{02}\mu_{21}\mu_{03} + \\ &\quad 9\mu_{20}^2\mu_{02}\mu_{12} + 12\mu_{20}\mu_{11}^2\mu_{21}\mu_{03} + 6\mu_{20}\mu_{11}\mu_{02}\mu_{30}\mu_{03} - \\ &\quad 18\mu_{20}\mu_{11}\mu_{02}\mu_{21}\mu_{12} - 8\mu_{11}^3\mu_{30}\mu_{03} - 6\mu_{20}\mu_{02}^2\mu_{30}\mu_{12} + \\ &\quad 9\mu_{20}\mu_{02}^2\mu_{21} + 12\mu_{11}^2\mu_{02}\mu_{30}\mu_{12} - 6\mu_{11}\mu_{02}^2\mu_{30}\mu_{21} \\ &\quad + \mu_{02}^3\mu_{30}^2)/\mu_{00}^{11} \end{aligned}\right\} (2)$$

VI. CLASSIFICATION

A. K-Nearest Neighbor Classifier

The k-nearest neighbor algorithm (k-NN) is a method for classifying objects based on closest training examples in the feature space. The k-nearest neighbor algorithm is amongst the simplest of all machine learning algorithms: an object is classified by a majority vote of its neighbors, with the object being assigned to the class most common amongst its k nearest neighbors (k is a positive integer, typically small). [52].

B. Principal Component Analysis

Principal Components Analysis (PCA) is a multivariate procedure, which rotates the data such that maximum variability's are projected onto the axes [53]. The main use of PCA is to reduce the dimensionality of a data set while retaining as much information as is possible. It computes a compact and optimal description of the data set. Data points are vectors in a multidimensional space. PCA is mathematically defined as an orthogonal linear transformation that transforms the data to a new coordinate system such that the greatest variance by any projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on [54-56]. PCA is theoretically the optimum transform for a given data in least square terms. The Principal Component Analysis module in proposed system generates a set of data, which can be used as features in building feature vector section. Fig. 4 shows a co-ordinate system (X_1, X_2). Choose a basis vector such that these vector points in the direction of max variance of the data, say (Y_1, Y_2), and can be expressed as

$$\left.\begin{aligned} Y_1 &= X_1 \cos\theta - X_2 \sin\theta \\ Y_2 &= X_1 \sin\theta + X_2 \cos\theta \end{aligned}\right\} \ldots\ldots\ldots\ldots\ldots\ldots\ldots..(3).$$

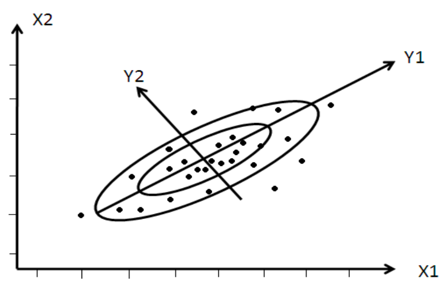


Fig. 4 Ellipse Distribution with PCA

We have used Euclidean distance as the similarity measure for comparing the testing and training samples

C. Gaussian distribution function

Number A membership function provides a measure of the degree of similarity of an element to a fuzzy set. Membership functions can take any form, but there are some common examples that appear in real applications. Membership functions can either be chosen by the user arbitrarily, based on the user's experience. This can also be designed using machine learning methods.

For an unknown input numeral x, the features are extracted using the affine invariant moments model. The membership function is chosen as,

$$\mu_i = \exp(x_i - M_i)^2 / 2\sigma_i^2 \ldots\ldots\ldots\ldots (4)$$

where, x_i is the i^{th} feature of the unknown numeral.

If all xi's are close to μ_i which represent the known statistics of a reference character, then the unknown numeral is identified with this known numeral because all membership function values are close to 1 and hence the average membership function is almost 1 [57].

Let, $M_i(r)$ and $\sigma_i^2(r)$ belong to the r^{th} reference numeral with r = 0,1…9, we then calculate the average membership as,

$$\mu_{av}(r) = 1/c \sum_{i=1}^{c} \exp(x_i - M_i)^2 / 2\sigma_i^2 \quad \text{....... (5)}$$

where c denotes for the number of fuzzy sets. Then x€r if $\mu_{av}(r)$ is the maximum for r=0,1…9.

D. *Support Vector Machine (SVM)*

A support vector machine (SVM) is a concept in computer science for a set of related supervised learning methods that analyze data and recognize patterns, used for classification and regression analysis. The standard SVM takes a set of input data and predicts, for each given input, which of two possible classes the input is a member of, which makes the SVM a non-probabilistic binary linear classifier. Support vector machine [58-60] is new classifier that is extensively used in many pattern recognition applications. On pattern classification problem, SVM demonstrate excellent generalization performance in practical applications. Support Vector Machines are based on the concept of decision planes that define decision boundaries. A decision plane is one that separates between a set of objects having different class memberships. Such functions are referred to as a kernel in the SVM approach. A kernel is utilized to map the input data to a higher dimensional feature space so that the problem becomes linearly separable. The kernel plays a very important role. Gaussian Radial basis function (kernel function) is given by-

$$K(x_i, x_j) = \exp\left(- \frac{\left\| x_i - x_j \right\|^2}{2\sigma^2} \right) \quad \text{........ (6)}$$

VII. EXPERIMENTAL RESULTS

Numerals 1, 2, 5, 6 and 7 have shown less recognition rate in case of PCA as compared to others as shown in Fig (5). Numerals 8, 4 and 0 show optimum results for the PCA. When we see the recognition rates of SVM from Fig 5, we find that SVM have shown good results for numerals 0,1,4,8, and 9 while other numerals show less rate of recognition but more better than that shown by PCA. As compared to PCA, SVM has shown improved results of 12% for numeral zero. Numeral one is being recognized 15.75% more by SVM than that by PCA[61].

Numeral two shows the difference of 9.75% in positive side for SVM as that for PCA. There is just a difference of approximately 4% for numerals three and four; when compared by both the classifiers. Likewise numerals five, six, seven and eight have shown approximate 7% enhancement in results through SVM classifier. Among all recognition rates numeral 0, 1 and 2 had highest recognition rate of 97%, 95% and 90% respectively for SVM. Numeral 3 have shown equal recognition rate for SVM and KNN [62] at 90% accuracy. For numerals 4 and 5 Gaussian has shown the spike at 96% and 93% respectively. For numerals 6 and 7, it is seen that they score comparatively less recognition with peaks at 86% and 89% respectively for SVM. Also as compared to overall recognition rate, SVM have shown recognition rate of 92.28% whereas PCA shows 84.1%, Gaussian shows 87.2% and KNN shows 90.04%. One can observe that SVM proves to be better classifier than Gaussian, PCA and KNN classifiers for affine invariant moments as feature extraction technique.

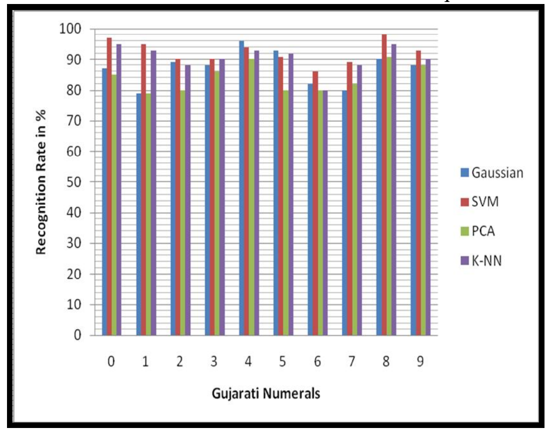


Fig. 5 Comparison of classifiers for affine invariant moments

CONCLUSION

As compared to overall recognition rate from Fig 5 SVM have shown recognition rate of 92.28%, Gaussian distribution function procured 87.2%. K-NN classifier has shown 90.04% whereas PCA shows 84.1%. One can observe from Fig. 5 that SVM proves to be better classifier than PCA, K-NN and Gaussian distribution function classifier for affine invariant moments as feature extraction technique. The results were compared with [9, 45-46] and were found to be better because we have applied our algorithm on noisy numerals. In future we will try to improve the recognition by doing the modification in the current system.

REFERENCES

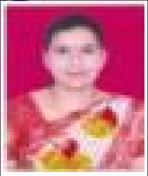
[1] 1. L. Harmon. "Automatic recognition of print and script" Proc. IEEE, 60, 1972, pp.1165-1176.
[2] Nafiz Arica, F.T. Yarman-Vural, (2001) "An Overview of Character Recognition Focused on Off-line Handwriting", *IEEE Trans. on System, Man, And Cybernetics – Part C: Application and Reviews*, Vol.31 No.2
[3] http//ccat.sas.upenn.edu
[4] http//languages.iloveindia.com
[5] http//india.mapsofindia.com
[6] http//en.wikipedia.org
[7] http://tdil.mit.gov.in/GujaratiScriptDetailsApr02.pdf Accessed on 10 January 2007
[8] S. Antanani and L. Agnihotri. "Gujarati character recognition". *Proc. of the 5th Int. Conf. on Document Analysis and Recognition (ICDAR)*, pages 418–421, 1999.
[9] Desai A. A.:(2010). "Gujarati handwritten numeral optical character reorganization through neural network", *Pattern Recognition*, Vol. 43, pp 2582-2589
[10] Dholakia J., Negi A., S. Rama Mohan (2005) "Zone Identification in the Printed Gujarati Text", *Proc. of 8th ICDAR*, p272-276
[11] Dholakia J., Yajnik A., Negi A.(2007) "Wavelet Feature Based Confusion Character Sets for Gujarati Script", *ICCIMA*, p366-371
[12] Veena Bansal (1999) Integrating knowledge sources in Devnagari text recognition PhD Thesis, IIT Kanpur.
[13] S. D. Connel, R.M.K. Sinha, A. K. Jain, "Recognition of Un-constrained On-Line Devnagari Characters". *Proceedings of ICPR*, pp. 2368 - 2371, Barcelona, Spain, September, 2000.
[14] R. M. K. Sinha and Veena Bansal, "A complete OCR for Printed Hindi Text in Devnagari Script". *Proceedings of Fifth International Conference on Document Analysis and Recognition*, pp. 800 - 804, Seattle, USA, September, 2001.

[15] V. Bansal and R. M. K. Sinha. "A complete OCR for printed Hindi text in Devanagari script". *In Proc. of the 6th Int. Conf. on Document Analysis and Recognition (ICDAR*), pages 800–804, 2001.
[16] B.B. Chaudhuri and U. Pal, "An OCR system to read two Indian Language Scripts: Bangla and Devnagari(Hindi)". *Proceedings of Fourth International Conference on Document Analysis and Recognition*, pp. 1011 - 1015, Ulm, Germany, August, 1997.
[17] Pal U and Chaudhuri B B (1997) "Printed Devnagari script OCR system". *Vivek* 10:12-24
[18] Lakshmi, C.V., Jain, R., Patvardhan, C. (2007) "Handwritten Devnagari Numerals Recognition with Higher Accuracy" *In International Conference on Computational Intelligence and Multimedia Applications:* Volume : 3 : 255
[19] Pal, U., Wakabayashi, T., Kimura, F. 2009. "Comparative Study of Devanagari Handwritten Character Recognition Using Different Feature and Classifiers" *10th International Conference on Document Analysis and Recognition, ICDAR '09* : 1111
[20] Anilkumar N. Holambe, Dr.Ravinder.C.Thool 2010 "Printed and Handwritten Character &Number Recognition of Devanagari Script using SVM and KNN" *International J. of Recent Trends in Engineering and Technology*, Vol. 3, No. 2, May 2010 163-166
[21] Rajashekararadhya, S.V. Vanaja Ranjan, P. (2009) "Zone-based hybrid feature extraction algorithm for handwritten numeral recognition of two popular Indian scripts", *Nature & Biologically Inspired Computing*, p 526 – 530
[22] Ragha L. R. Sasikumar M. 2010 "Using moments features from Gabor directional images for Kannada handwriting character recognition" *Proceedings of the International Conference and Workshop on Emerging Trends in Technology*
[23] Sharma D. and Jhajj P. (2010) "Recognition of Isolated Handwritten Characters in Gurmukhi Script" *International Journal of Computer Applications* 4(8):9-17.
[24] Antarpreet Kaur , Rajiv K. Sharma, and Amardeep Singh 2010 "A Hybrid Approach to Classify Gurmukhi Script Characters" *International J. of Recent Trends in Engineering and Technology*, Vol. 3, No. 2, May 2010 103-105
[25] Singh K. S. Jangid M., Dhir R. Rajneesh Rani,(2011) "Handwritten Gurmukhi Character Recognition Using Statistical and Background Directional Distribution Features" *International Journal on Computer Science and Engineering* 3(6): 2332-2345.
[26] Journal of Language Technology, Vishwabharat@tdil, July2003.
[27] Journal of Language Technology, Vishwabharat@tdil,July,2004.
[28] Chaudhari B B, Pal U and Mitra M (2002) "Automatic Recognition of printed Oriya script." *Sadhana* 27(1): 23-34.
[29] Mohanty S Behera H K (2004) "A Complete OCR Development System for Oriya Script". *Proceeding of Symposium on Indian Morphology, phonology and Language Engineering*
[30] Pal, U. Wakabayashi, T. Sharma, N. Kimura, F. (2007) "Handwritten Numeral Recognition of Six Popular Indian Scripts." Proc ICDAR, 2: 749 – 753.
[31] Soumya Mishra, Debashish Nanda , Sanghamitra Mohanty 2010 "Oriya Character Recognition using Neural Networks" *Special Issue of IJCCT Vol. 2 Issue 2, 3, 4; 2010 for International Conference [ICCT-2010]*, 88-92.
[32] Bhattacharya, U. Ghosh, S.K. Parui, S.K. (2007) "A Two Stage Recognition Scheme for Handwritten Tamil Characters" *Proc. ICDAR*, 1:511 – 515.
[33] Sarveswaran, K. Ratnaweera, D.A.A.C (2007) "An adaptive technique for handwritten Tamil character recognition", *Proc. ICIAS*, p151 - 156
[34] Bremananth, R. Prakash, A. (2009) "Tamil Numerals Identification", *Proc.International Conference on ARTCom*, p 620 - 622
[35] Ramanathan, R. Ponmathavan, S. L.Thaneshwaran, N. Valliappan Nair, Arun S. Soman, K.P. (2009) "Optical Character Recognition for English and Tamil Using Support Vector Machines", *Proc. International Conference on ACT '09*, p 610 – 612.
[36] R.Ramanathan, S.Ponmathavan, L.Thaneshwaran, Arun.S.Nair, N.Valliappan and Dr. K.P.Soman (2009) "Tamil Font Recognition Using Gabor Filters and Support Vector Machines", *International Conference on Advances in Computing, Control, and Telecommunication Technologies*, pp 613-615.
[37] Rajashekararadhya, S.V. Ranjan, P.V (2010) "The Zone-Based Projection Distance Feature Extraction Method for Handwritten Numeral/Mixed Numerals Recognition of Indian Scripts" *Proc ICFHR*, p 617 – 622.
[38] Jawahar C. V., M. N. S. S. K. Pavan Kumar, S. S. Ravi Kiran 2003. A Bilingual OCR for Hindi-Telugu Documents and its Applications. *International Conference on Document Analysis and Recognition1*:408-412
[39] Lakshmi C V , C Patvardhan, 2003 Optical Character Recognition of Basic Symbols in Printed Telugu Text. *IE(I)Journal*-CP Vol 84, pgs.66-71.
[40] Pratap Reddy L., Satyaprasad L., A.S.C.S. Sastry, "Middle Zone Component Extraction and Recognition of Telugu Document Image," icdar, vol. 2, pp.584-588, *Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)* Vol 2, 2007
[41] Anuradha Srinivas, Arun Agarwal, C.R.Rao 2007 Telugu Character Recognition. *Proc. Of International conference on systemics, cybernetics, and informatics*, Hyderabad, pgs.654-659
[42] Lawrance Lo. at http://www.ancientscripts.com/ Accessed 15 June 2007
[43] Kayasth M and Patel B (2009) "Offline typed Gujarati Character recognition". *NJSIT* pp73-82
[44] Chaudhari S and Gulati R M (2010) "A font size independent OCR for machine printed Gujarati numerals". 3(1):70-78
[45] Prasad, J.R. ; Kulkarni, U.V. ; Prasad, R.S. (2009) "Template Matching Algorithm for Gujarati Character Recognition" *In Proc. Of 2nd International Conference on Emerging Trends in Engineering and Technology (ICETET)*, pp 263 - 268
[46] Prasad, J.R.; Kulkarni, U.V.; Prasad, R.S. (2009) "Offline Handwritten Character Recognition of Gujarati script using Pattern Matching" *In Proc. Of 3rd International Conference on Anti-counterfeiting, Security, and Identification in Communication*, pp 611 - 615
[47] E. Hassan, S. Chaudhury, M. Gopal, and J. Dholakia, "Use of MKL as symbol classifier for Gujarati character recognition", *in Proc. Document Analysis Systems*, 2010, pp.255-262.
[48] Desai A. A. (2010) "Handwritten Gujarati Numeral Optical Character Recognition using Hybrid Feature Extraction Technique". *IPCV* pp 733-739
[49] T. Wakahara, K. Odaka, "Adaptive Normalization of handwritten Characters using Global/Local Affine Transformation", *4th Int. Conference on Document Analysis and Recognition, (ICDAR'97)*, 1997.
[50] Jan Flusser, Tomas Suk, "Affine Moment Invariants: A new tool for Character Recognition", *Pattern Recognition Letters*, vol.15, pp.433-436, 1994
[51] Rahtu E., Salo M., Heikkila J. and Flusser J. (2006) *18th International Conference on Pattern Recognition, Published by IEEE Computer Society in IEEE Xplore USA*,634-637
[52] Hall P., Park B.U., Samworth R.J. "Choice of neighbor order in nearest neighbor classification", *Annals of Statistics* vol 36 (5): 2135-2152 2008.
[53] Ramteke R.J., Borkar P.D. and Mehrotra; "Recognition of isolated Marathi Handwritten Numerals: An Invariant Moments Approach" *Proceedings of the International Conference on Cognition and Recognition 2005*, pp 482-489
[54] Smith L. I.. "A Tutorial on Principal Components Analyzing – Discussion and singular value decomposition", http://www.cs.otago.ac.nz/cosc453/ student_tutorials/principal_components.pdf (2002).
[55] Shlens J. "A Tutorial on Principal Component Analysis", available at: http://www.cs.princeton.edu/picasso/mats/PCA-Tutorial-Intuition_jp.pdf (2005).
[56] Jolliffe I. T. "Principal Component Analysis", Springer Series in Statistics, 2nd ed., Springer. (2002)..

s

[57] Hanmandlu M and Murthy OV R, (2005) "Fuzzy Model Based Recognition of Handwritten Hindi Numerals". *Proc ICCR*, p 490-496
[58] V. Vapnik, "The Nature of Statistical Learning Theory" *Springer Verlag*, 1994.
[59] C. Burges, "A Tutorial on support Vector machines for pattern recognition" *Data mining and knowledge discovery*, vol.2,1998,pp.1-43.
[60] V.N. Vapnik, Statistical Learning Theory. *John Wiley and sons*, 1998
[61] Baheti M.J., Mane A.V., Hannan M.S. and Kale K.V.(2011) "Comparison of PCA and SVM for a West Indian Script-Gujarati" *CiiT International Journal of Digital Image Processing,* 3(11), 709-715.
[62] Baheti M. J., Kale K.V., Jadhav M.E, "Comparison Of Classifiers For Gujarati Numeral Recognition" *International Journal of Machine Intelligence*, Volume 3, Issue 3, 2011, pp-160-163

AUTHOR'S PROFILE



Mamta Baheti

She has received M.Sc. Degree in Computer Science from Sant Gadge Baba Amravati University in 2003. Since 2007, she has been a Ph.D. student of Dr. K. V. Kale. Her current research interests include pattern recognition, image analysis and document processing. She is presently working as Lecturer at Department of Computer Engineering, in SNJB's College of Engineering, Chandwad, Nashik.

Dr. Karbhari Vishwanath Kale

He is M.Sc., MCA (Engg & Tech), Ph.D., FIETE., Professor and Head Dept. of Computer Science and Information Technology, Dr. Babasaheb Ambedkar Marathwada, University, Aurangabad, INDIA Completed M.Sc Physics (Electronics) in 1987, B.Ed in 1989, MCA (Engg. And Tech) in 1995 and Ph.D on Superionics in 1997. Seventeen students awarded Ph. D. in Computer Science. Ten students have working for Ph.D. under the guidance. One student awarded and two are working for M.Phil. under the guidance . He is IEEE fellow member