A Review of key features instrumental in recognizing the Multi-lingual Optical Characters

Rohit Verma, Research Scholar, IKG Punjab Technical University, Jalandhar,India,
rohitverma11@gmail.com
Dr. Jahid Ali,Director, Shri Sai Iqbal College of Management & IT,Pathankot, India,
zahidsabri@rediff.com

Abstract - In the modern era extremely deep research is going on in the field of Optical Character Recognition. The most challenging part of the research is generation of feature vector to identify characters. Significantly important features can recognize the characters with great accuracy. This paper summarize numerous feature extraction techniques that would be greatly helpful for the researchers to carry out their work with the diversity of Character Recognition Techniques. Considerably important features are independent of font style and invariant of scalability, orientation and translation. Decades have been devoted to develop a highly accurate and complete feature vector that could recognize the characters with high precision just like humans and present fantastic results. Most of the feature vectors generate great results with a few font styles and characters. But still a challenging task is there to recognize multi-lingual characters simultaneously and separately in a single document. Therefore it is believed that still scope is there to work in this field for generating features that could accurately and efficiently recognize multi-lingual characters simultaneously with high precision.

Keywords-- Optical Character Recognition, Feature Extraction, Patter Recognition, Image Processing, OCR, image processing.

I. INTRODUCTION

Optical Character Recognition (OCR) describes task of identification of characters and numerals from the scanned images and converted them to the modifiable form for further processing. The earliest research work on CR was reported in early 1870 when an image transmission system termed as retina scanner using a collection of photocells was invented by Carey[1].

In the beginning phases OCR was developed to produce reading aids for blinds. But now the spectrum of CR (Character Recognition) applications become very vast. Myriads of applications are using the concept of Character Recognition[2]. Though present research on CR is prove to be fruitful in case of recognition of bilingual and multilingual printed character with more than 95% accuracy[3]. Even a lot of research work has been reported on hand written characters with considerably high accuracy. But most of work is done on western countries script. Therefore, there is a huge scope of research work on identification of features for multilingual and Indian scripts. Nearly twelve different scripts are used to write official languages[4] of India.

The objective of Character recognition can be achieved by the three extremely important steps: Segmentation, Feature Extraction and Classification. Out of these the most challenging one is feature extraction. The term feature extraction refers to the process of collecting extremely important character shape related relevant information so that vital objective of classifying characters can be achieved by a formal procedure. This paper describes various feature extraction techniques that would be helpful for the researchers to work in the field of Character Recognition. **Section 1** is just the introductory part of the OCR system , **Section 2** classify and describe extremely important features that could be considered for the characters recognition system as well as presented the results of experiments conducted in the field of CR(Character Recognition) whereas **Section 3** Depicts the performance comparisons recorded for various features on the basis of literature.

II. FEATURE CLASSIFICATION FOR CHARACTER RECOGNITION

A feature is defined as “image pattern which differs from its immediate neighbors” [5]. Some of the extremely important characteristics of feature extracting algorithms are: Robustness, Accuracy, Repeatability, Generality, Efficiency and quantity[6].

 DOI : 10.35291/2454-9150.2019.0088

Table 1. Depicts the classification of prominent features for character recognition, their sub categories and important characteristics.

Techniques Available for Feature Extraction		
Shape Oriented Features		
		Important Characteristics
Moment related Features	Hu's Invariant Moments	• Any character is represented by density distribution function[5] • Zernike moments are invariant to rotation and because of orthogonal basis, they show minimum information redundancy[7]. • Hu's Moments are Translation, Rotation and Scale invariant [8] • The shape and contour of the input image is analyzed for recognition purpose.
	Invariant Moments	
	Zernike Moments	
Native features	Horizontal, Vertical and oblique line types	• Here horizontal, vertical and oblique lines that form the character are analyzed for the recognition of the characters. • The eccentricity reports the degree of deviation from being circle. • Extent is the ratio of number of pixels in the region to the pixels in the bounding box.
	Different types of arcs	
	Different types of loops	
	Eccentricity	
	Extent	
Geometrical Features	Angular width Calculation	• Geometric features are constructed by a set of fundamental geometric elements like points, lines, arcs, curves, surfaces, Blobs etc. • Beam Angle Statics (BAS) is a shape descriptor which is scale, rotation and translation invariant[9].
	Area Calculation	
	Width Features	
	Geometrical distances	
Spatial Relation Features	Basic Chain Code	• Spatial Relation pixels are generated from contour pixels. • Accuracy of 94% is reported with English and numerals using chain code Native features [10]. • Chain code techniques preserve the shape information of the character and contribute to the data reduction. • Vertex Chain Code gives fantastic results for shapes having rectangular, triangular and hexagonal cell structures. • VCC is also invariant under translation and rotation[11].
	Chain Code Histogram	
	Vertex Chain code (VCC)	
	Extended Vertex Chain Code	
	Variable vertex Chain Code(VCC)	
	Complex Vertex Chain Code	
	Topology Based Features	
	Regular Expression	

Non-Shape Based Features	
Feature Name	**Important Characteristics**
Centroids	• Centroid is the Arithmetic mean position of all the pixels contributing the shape of the image. • Centroid of an image is allocated near the high density area of the image[12]. • Combination of Template matching technique like Normalized Cross Correlation or Dice coefficient with suitable decision making algorithm could generate high accuracy model [13] for image recognition.
Spatial Dots	
End Points	
Intersection Points	
Normalized Cross Correlation	
Template Matching	
Dice Coefficient	
Statistical Features	
Crossing Corners	• Skewness provides the degree of asymmetry of gray levels around the sample mean.
Crossing Counts	

 DOI : 10.35291/2454-9150.2019.0088

Features related to density	• Kurtosis provides the shape of tail of the histogram. • Statistical features are suitable if the texture primitives are comparable to those of pixels. • Probability distribution describes the distribution of pixel values to define an image pattern. • Extent defines the ratio of the area of the character to the area of the bounding box in which character is enclosed. Recognition with extent is not a appreciable approximation. • Crossing corners and crossing counts are the prominent features to infer the contents of an image.
Features related to longest run length	
Peak extent features	
Mean, Standard Deviation, Variance, Skewness, Kurtosis	
Euler Number(Number of holes in an image)	
Hole parameters	
Holes' position	
Probability Distribution	
Projections	
Profiles	
Shadow related Features	

Author's compilation from different sources

Broadly speaking, there are two types of features: Shape related features and non-shape depending features. Shape related features should be equipped with certain vital characteristics: identifiability, noise resistance, reliability and translation, rotation as well as scale invariance for instance. Non-shape depending features include statistical information of the characters, number of holes, projections, intersection, template matching and many more.

A Shape-Based features

These features should meet the requirements of completeness, compactness, simplicity, stability, large scope and uniqueness[14]. Statistical Independence, occultation resistance, identifiability are the essential properties of Shape-Based features. Occultation resistance means when a part of the object is concealed because of some reason, the features of the remaining part must not change. Shape-dependent features are divided into four sub-categories viz. Moment related features, Native features, Geometrical features and spatial relation features.

Moment related Features

Moment-related features are proposed to be pattern sensitive features in the field classification and recognition applications. Moments based features analyze the distribution of various pixels in the image relative to a fixed point. Since distribution of various pixels relative to a fixed point in the image is invariant under rotation, scaling or translations. So better results of recognition could be expected. This section presents a few papers from literature which have made use of Zernike moments, Hu's invariant moments and invariant moment of character recognition.

A general expression for regular moment of order (p + q) can be represented by the moment function

$$m_{pq} = \int \int x^p y^q f(x, y)\, dx\, dy$$

Here f(x, y) is the image intensity function.

For a digital image, translation invariant moment is given by

$$\mu_{pq} = \sum\sum (x - xc)^p (y - yc)^p f(x, y)$$

Here, $xc = m_{10}/m_{00}$ and $yc = m_{01}/m_{00}$ are the co-ordinates of the centroid.

Paper [15] describes an approach based on modified invariant moments for character recognition in multi-font English characters. B. V. Dhandra et al. reported an accuracy of 99.75% with 3640 Capital letters English characters and 99.56% with 3640 small letters English Characters [15]. In paper [16] Kale et al. reported average accuracy of 98.37% and 95.82% with basic devnagari characters using Zernike moments. The character image to be recognized is splitted into nine zones of equal size and then zernike moments based features are extracted from these zones. These features were applied to Support Vector Machine and K-Nearest Neighbour for classification and testing purpose. Successful result of more than 95% accuracy was reported. In case of compound devnagari characters this paper reported an average accuracy of 98.32% and 95.42% with SVM and K-NN respectively.

Dan L. Lacrama reported an accuracy of 93.75% in paper [17] by considering mixed moment sets Hu's, Zernike and Affine moments.

Native Features

Native features include various line types, their orientations, arcs present in the characters, number of curves and their types, number of loops etc. These features can be obtained by thinning the character input image to one pixel. In paper [18] Raja and John extracted horizontal and vertical lines with respect to the centroid from Tamil characters and termed as HLAC for horizontal line above centroid and HLBC for horizontal line below centroid. Similarly vertical lines are classified as VLRC and VLLC for vertical line to the left side of the centroid and vertical line to the right side of centroid. The input character is then dig for curves and these curves were classified as curve-1, curve-2, curve-3, curve-4. Afterwards small loops and big loops were extracted using chain codes and classified as small loops above centroid, small loop (SL) below centroid, big loop above centroid (BL) and big loop (BL) below centroid. These 12 extracted features were applied to BPNN, SVM and DT classifiers as inputs. 95.09% recognition was reported using novel hybrid decision tree. A new method of recognition of hand written Chinese characters was proposed by Lu et al. in the paper [19]. A structural representation called hierarchical attributed graph representation (HAGR) was proposed and an accuracy of 90.45% was reported.

 DOI : 10.35291/2454-9150.2019.0088

Spatial Relational Features

These features are extracted by the analysis of neighboring pixels in the image as they use location/spatial information of the image. Local Binary Pattern (LBP), Chain Code, Vertex Chain Code(VCC), Topology based features, variable VCC, regular expression and chain code histogram are some of the instances of spatial relational features that can be used for efficiently recognition of images. H. Tirandaz et al. in paper [20] proposed a new approach based on morphological structure for Arabic characters identification. Their method is based on center of mass which is invariant under scaling, orientation and translation. In addition, topology based features like intersection pixels, masking the intersection of loops and ends points are considered for recognition. An accuracy of 92% was reported by the method proposed. In paper [21], Sebastian et al. proposed a novel-frame work for object recognition using the concept of edit shock graph of 2D shapes. According to the proposed approach, the given shape is divided into equivalent classes. The shapes having similar shock graph are considered to be equivalent. Accuracy rate of 100% is reported in top three matches with two distinct databases of 99 and 216 shapes. In paper [22] a hybrid approach for classification of Devnagari characters using minimum edit distance, regular expression and chain code features, proposed by Deshpande et al. The given character was converted into a chain code and it is matched with regular expression. Minimum edit distance classifier is then applied to unmatched results. Using samples of 5000 dataset items accuracy of 82% was reported. Liu et al. in paper [23] proposed three updated and improved version of chain codes: Extended Vertex chain code(E_VCC), variable length vertex chain code (V_VCC) and the last one is compressed vertex chain code (C_VCC). The proposed compressed vertex chain code is based on Huffman method. From the literature it is reported that C_VCC is the most efficient and it uses lesser number of bits as compared to others.

Geometrical Features

Geometric features are composed of set of fundamental geometric primitives like points, curves, lines and arcs or surfaces, for instance. They include edge features, corner features, Blobs, perimeter calculations, area calculations, angular width and perpendicular distances etc. Vikas J. Dongre and et al. in the paper [24] proposed an approach to recognize Devnagari numerals using seventeen geometric and structural features. It is reported that combination of classifiers offer better results over individual classifiers. Average accuracy of 81.67% is reported for 1500 images of Devnagari numerals. In paper [9] Arica and Yarman Vural proposed shape descriptor based on beams originated from a boundary point. Beam is the line obtained by joining a point on the boundary to the rest of the points on the boundary. Then angle between a pair of beams is calculated to extract the topological structure of the boundary. A shape descriptor using third order statistics called Beam Angle Statistics (BAS) of all beam angles was proposed. It must be noticed that BAS is invariant to rotation, scaling and translation. In paper [25] , Iqbal et al. proposed a novel algorithm called Radial Sector Coding (RSC) to extract features which are translational, scale and rotation invariant. Each character image is divided into sectors considering centroid as center to the boundary points called cut points. Normalized distances are calculated to the cut points using maximum distance. 98.6% accuracy was reported by using Multilayer Feed Forward Artificial Neural Network (ANN). Asaari and Rosdi in paper [26] proposed a new geometric for the infrared single finger image recognition system. Two features: finger widths and fingertip angles after extraction are transformed into frequency domain using Discrete Fourier Transformation (DFT) to make them translation and rotation invariant. The new feature sets obtained from DFT are blended by concatenating them into a single row vector called “width and fingertip angle”. Significantly high success rate was found in the proposed method. In paper [27], Pal and Tripathy proposed a novel scheme for the recognition of multi-sized and multi-oriented isolated characters of printed scripts. Distances between the centroid and outer contours of the characters are calculated and re-arranged to achieve size and rotation invariant features. Success rate of 97.8% and 98.1% was reported for Bangla and Devnagari characters.

B Non Shape-Based features

Non shape-based features recognize the images in terms of mathematical values. These features may include centroid of the character image, Intersection points, end-points, normalized cross correlation, dice coefficient, template matching etc. Statistical features for instance mean, standard deviation, kurtosis, hole number, location of hole, probability distribution, projection, profiles, density based features all comes under the category of statistical features. Some of these features are used in combination with other features for describing a character accurately. S. Abirami et al. in paper [28] proposed six statistical features extracted from character boundaries to classify the Tamil hand written characters using symbol modeling HMM. Ten different samples for each of the 60 Tamil characters were collected to perform the experiment and result with 85% accuracy was reported. In paper [29], Swapnil Desai et al. proposed a method for character recognition that used template matching approach. It is observed that template matching when applied with ANN accuracy of 91.82% was reported. Nazanin Sadat et al. in paper [13] presented mathematical basis of template matching techniques and various matching algorithms. It is also proposed that template matching should be accompanied with post classification or decision making step in order to achieve the high degree of accuracy. Furthermore, it is

 DOI : 10.35291/2454-9150.2019.0088

added that feature based methods are applied when structural information instead of intensity information is available and template matching is always applied in combination with other algorithms so that desirable results could be achieved.

Composite/Combined Features

Composite features may be defined as the coupling of shape-oriented and non-shape oriented features. Generally to increase the degree of accuracy, both of these features are used in combination. In this section, few papers from literature which are using the composition of shape oriented and non-oriented features are discussed. A character recognition system for Kannada, Malayalam, Tamil ,Telugu and English printed characters was proposed by Manjunath Aradhya et al. [30] Principal Component Analysis (PCA) and Fourier transform which are the classical feature extraction and data representation techniques used in the proposed method. Average accuracy of 95.1% recognition was reported with 137,500 character dataset. N.R Soora et al. proposed two novel geometrical features extraction techniques for English letters and numerals. Two types of geometrical features and a statistical crossing count feature were extracted with the help of horizontal, vertical, right diagonal, and left diagonal scan lines and centroid of the character to be recognized. The first feature vector computed the global shape of the input character by computing the angle between the extreme end points and the second feature vector is developed by finding the angle between the lines joining the centroid of the given character to the points of Intersection of the scan line touching the border of the character. Accuracy of 98.8% was reported by them in the proposed method with 741 media-lab bench mark database.

Kahan et al. in paper [31] proposed a composite approach called Line Adjacency Graph(LAG). In the proposed method several techniques were combined to improve the recognition rate of printed roman alphabets of any font and size. With the help of LAG-traversal approach for thinning, myriads of binary features like: number of holes, expected size and location of holes, concavities in skeleton structure, strokes crossing, for instance could be generated. These binary features then applied to statistical Bayesian classifier and recognition of 97% was reported in case of six different fonts and 99% in case of single font. In paper [32] Ping and Liuhi proposed classification system composed of Neural Network and Decision tree classifiers to recognize hand written numerals. In this hybrid classification system geometrical features of a characters are elaborated in details. After developing and extracting global features, a novel floating feature detector was developed. Global features included were width features, middle line feature, concave features and point features .The given work was tested with two data sets comprises of 7500 characters. Success rate of 97.85% was reported. Garain and Chaudhuri in the paper [33] presented an approach for online understanding of mathematical expression. The proposed method comprises of two extremely important stages: symbol recognition and structural analysis. To identify spatial relationships among symbols, myriads of online and offline features were used in structural analysis phase. For getting the high degree of accuracy two different classifiers were used. For the recognition of symbols the proposed method used neuromotor feature of handwriting. The experiments were conducted with a dataset of expressions count to 5500 expressions and accuracy of 74.92% was reported. Arica and Yarman Vural presented a review of offline handwritten character recognition approaches from the literature in paper [34].

Firstly the historical views of Character Recognition were presented, then available Character Recognition approaches with their strengths and weakness were reviewed and finally the current status was discussed for the researchers in the field. Four extremely vital approaches for characters Recognition: Template matching, Statistical techniques, structural analysis and Neural Network with their superiorities and weaknesses were unveiled for the researchers.

Table 2: Performance comparison of composite methods with accuracy[35]

S.No	Features	Test Data	Size of Test Data	Accuracy	Reference
1	Composite features	E, H, and Bengali	words : 4800	75%	[36]
		E (CAS Database)	characters: 3179	75.3%	[37]
		(CEDAR DATABASE)	--	80%	[38]
		E (BAC Database)	characters: 2183	84.57%	[40]
		D	characters: 1600	86.5%	[42]
		H and E	characters: 6600	95.4%	[53]
		E, Chinese & N	characters: 10,000	95.41%	[54]
		N, E Kana and Chinese	characters: 8493	97.88%	[63]
		Chinese, Japanese, N	---	92%	[50]
2	Native features as well as Neuromotor characteristics	Mathematical symbols	mathematical expressions: 5500	74.92%	[33]
3	Native features	B	characters: 23550	97.06%	[62]
	Native features	N	characters: 52466	99.4%	[51]
4	Native features & centroid features	Tamil	characters: 1000	100%	[43]
5	Crossing corners with left and right diagonal scan lines	E	Characters: 650	84.52%	[39]
6	updated trellis diagrams, Hamming distance and Viterbi features	Kannada	Characters: 650	85%	[41]

 DOI : 10.35291/2454-9150.2019.0088

7	--	E & N	images: 1287	89.1%	[43]
8	moment based lower and upper profiles, Density of pixels	D N	Characters: 2460	89.68%	[44]
9	Composite features(vertex points, moments, density of pixels & horizontal Zero Crossing)	D	--	90%	[45]
10	Shadow features and chain code histogram	D	Characters: 2460	90.74%	[46]
11	Improved projection profiles, translational statistical features	E	Characters: 2460	91.38%	[47]
12	Patter of foreground as well as background pixels in angular and radial directions	Kannada	pages: 16	91.56%	[48]
13	Template matching	Telugu	Characters: 2524	92%	[49]
14	--	Digits	Characters: 11,791	94.7%	[51]
15	Wavelet Features	Kannada	--	95%	[52]
16	--	Kannada, Telugu, Malayalam & E	characters: 137,500	95.1%	[30]
17	Crossing counts and peripheral background area ,Contours directional counts.	E and N	Characters: 1248	95.7%	[55]
18	Composite features	Arabic numerals	characters: 1000	95.80%	[43]
19	Strokes, run number based, water overflow for reservoir features	Oriya	--	96.3%	[56]
20	Octant, centroid and longest run length features as well as shadow features	B N	characters: 2000	96.67%	[57]
21	Variable gray levels and projection profiles	E & N	characters: 400	97.2%	[58]
22	Moments, projects, profiles and Native features	E	characters: 5274	97.4 %	[59]
23	--	Chinese	characters: 13053	97.4%	[60]
24	Ratio of foreground pixels	E and N	images: 805	97.52%	[61]
25	Local and global features	N	characters: 7500	97.85%	[32]
26	Horizontal & vertical projection histogram	E and N	characters: 241	97.9%	[64]
27	Class conditional probabilities and chain code histograms	B	characters: 4000	98.05%	[65]
28	Chain code and gray scale projections	N	characters: 6400	98.1%	[66]
29	CCs triangle area calculation and perpendicular distance features	D, Marathi and E	characters: 30000	98.5%	[3]
30	Projection and Chain codes	E and N	images: 10000	98.51%	[54]
31	Statistical and spatial relational features	E and N	characters: 196000	98.75%	[31]
32	Chain code, angular width and geometrical features	E and N	characters: 6584	98.8%	[37]
33	CCs, Perpendicular distances and triangle area calculation features	E and N	characters: 6584	99.03%	[3]
34	Statistical features, Structural and template matching approach	B	--	99.10%	[67]
35	Angular information obtained from external and internal contour of pixels	B and D	characters: 7515	99.18%	[27]
36	Contour & Structure encoding feature	Numerals	characters: 77,702	99.4%	[68]
37	Row and column nanograms	English	characters: 4084	100%	[69]

Note: E-English, H-Hindi, B-Bangla, D-Devanagari, H-Hindi, N-Numerals
Author's compilation

CONCLUSION

In this paper, extremely important features that can be considered for attaining high degree of character recognition accuracy are reviewed from the literature. On the basis of the approaches used, these features are classified as Shape-oriented features, non-shape oriented features and composite features. Some of the important characteristics of these features were also presented to understand the vital parameters for getting desirable results. Composite features also termed as hybrid features which are used in combination with other algorithms and their success rate of recognition are reported from the literature.

FUTURE WORK

From the literature it is found that not much research is submitted in the field of printed and handwritten Indian scripts in combination with Roman characters. Though, much effort is done in the recognition of individual scripts: Devnagari and Roman characters for instance, but printed Devnagari scripts and Roman characters in combination does not provide desirable recognition results. There are a few OCR systems that could recognize text containing Devnagari and Roman printed characters simultaneously with high degree of accuracy, but still some gaps are there to be worked upon.

REFERENCES

[1] J. Mantas, "An overview of character recognition methodologies," *Pattern Recognit.*, vol. 19, no. 6, pp. 425–430, 1986.

[2] A. Mir, A. Mir, S. A. Hannan, and Y. Perwej, "AN OVERVIEW AND APPLICATIONS OF OPTICAL CHARACTER RECOGNITION," *Int. J. Adv. Res. Sci. Eng.*, vol. 8354, no. 3, pp. 261–274, 2014.

[3] N. R. Soora and P. S. Deshpande, *Novel Geometrical Shape Feature Extraction Techniques for Multilingual Character Recognition*, vol. 4602, no. October. Taylor & Francis, 2016.

[4] U. Pal and B. B. Chaudhuri, "Indian script character recognition : a survey," vol. 37, pp. 1887–1899, 2004.

 DOI : 10.35291/2454-9150.2019.0088

[5] T. Tuytelaars and K. Mikolajczyk, "Local Invariant Feature Detectors : A Survey," vol. 3, no. 3, pp. 177–280, 2007.

[6] M. Hassaballah, A. A. Abdelmgeid, and H. A. Alshazly, "Image Features Detection , Description and Matching," *Springer*, pp. 11–46, 2016.

[7] R. S. Kunte and R. D. S. Samuel, "A simple and efficient optical character recognition system for basic symbols in printed Kannada text," *Sadhana 32 (5), pp. 521–533.*, vol. 32, no. October, pp. 521–533, 2007.

[8] M. K. Hu, "Visual Pattern Recognition by Moment Invariants," pp. 66–70, 1962.

[9] N. Arica and F. T. Y. Vural, "BAS : a perceptual shape descriptor based on the beam angle statistics," vol. 24, pp. 1627–1639, 2003.

[10] G. Hsu, J. Chen, and Y. Chung, "Application-Oriented License Plate Recognition," vol. 62, no. February 2013, pp. 552–561, 2013.

[11] A. M. Salem, A. A. Sewisy, and U. A. Elyan, "A Vertex Chain Code Approach for Image Recognition Boundary extraction," *ICGST-GVIP J.*, vol. 5, no. March, pp. 1–8, 2005.

[12] H. A. R. Khairul Hamimah ABAS, "S e n s o r s & T r a n s d u c e r s Thermal and Visible Sensor Application : Physiological Thermal Moment Invariant Analysis for Infrared-Based Face Identification," vol. 163, no. 1, pp. 262–269, 2014.

[13] N. Sadat, R. Babaei, A. Sadat, and B. Ghiasi, "Template Matching Advances and Applications in Image Analysis," *Am. Sci. Res. J. Eng. Technol. Sci.*, pp. 91–108, 2016.

[14] M. Yang, K. Kpalma, J. Ronsin, M. Yang, K. Kpalma, and J. Ronsin, "Shape-based invariant features extraction for object recognition," 2011.

[15] B. V. Dhandra, V. S. Malemath, H. Mallikarjun, and R. Hegadi, "Multi-font English character recognition based on modified invariant moments," *J. Comb. Math. Comb. Comput.*, vol. 67, pp. 153–162, 2008.

[16] K. V Kale, S. Member, P. D. Deshmukh, S. V Chavan, and S. Member, "Zernike Moment Feature Extraction for Handwritten Devanagari (Marathi) Compound Character Recognition," vol. 3, no. 1, pp. 68–76, 2014.

[17] Dan L. Lacrama, "The use of invariant moments in hand-written character recognition," 2006.

[18] P. Taylor, S. Raja, and M. John, "Tree Classifier A Novel Tamil Character Recognition Using Decision Tree Classifier," no. October, pp. 37–41, 2014.

[19] S. I. W. E. I. Lu, "H i e r a r c h i c a l attributed g r a p h representation a n d recognition of h a n d w r i t t e n chinese c h a r a c t e r s," vol. 24, no. 7, pp. 617–632, 1991.

[20] H. Tirandaz, M. Ahmadnia, and H. Tavakoli, "GEOMETRIC-TOPOLOGICAL BASED ARABIC CHARACTER RECOGNITION , A NEW APPROACH," vol. 95, no. 15, 2017.

[21] T. B. Sebastian, P. N. Klein, and B. B. Kimia, "Recognition of Shapes by Editing Their Shock Graphs," vol. 26, no. 5, pp. 550–571, 2004.

[22] L. M. and S. A. Parag S. Deshpande, "Fine Classification & Recognition of Hand Written Devnagari Characters with Regular Expressions & Minimum Edit Distance Method," *J. Comput.*, vol. VOL. 3, NO, no. May, pp. 10–17, 2008.

[23] Y. Kui, W. Wei, P. Jie, and B. Žalik, "Compressed vertex chain codes," vol. 40, pp. 2908–2913, 2007.

[24] V. J. Dongre, "Devnagari Handwritten Numeral Recognition using Geometric Features and Statistical combination Classifier," *Int. J. Comput. Sci. Eng.*, vol. 5, no. 10, pp. 856–863, 2013.

[25] A. Iqbal, A. B. M. Musa, A. Tahsin, A. Sattar, M. Islam, and K. Murase, "A Novel Algorithm for Translation , Rotation and Scale Invariant Character Recognition," *IRE Trans. Inform. Theory*, pp. 1367–1372, 1962.

[26] M. Shahrimie, M. Asaari, and N. Tebal, "A Single Finger Geometry Recognition Based On Widths And Fingertip Angles (WFTA)," pp. 1–4, 2013.

[27] U. Pal and N. Tripathy, "A contour distance-based approach for multi-oriented and multi-sized character recognition," vol. 34, no. October, pp. 755–765, 2009.

[28] S. Abirami, V. Essakiammal, and R. Baskaran, "Statistical features based character recognition for offline handwritten Tamil document images using HMM," *Int. J. Comput. Vis. Robot.*, vol. 5, no. 4, pp. 422–440, 2015.

[29] A. S. Swapnil Desai, "Optical character recognition using template matching and back propagation algorithm," *Int. Conf. Inven. Comput. Technol.*, 2016.

[30] V. N. M. A. Ã, G. H. Kumar, and S. Noushath, "Multilingual OCR system for South Indian scripts and English documents : An approach based on Fourier transform and principal component analysis," vol. 21, pp. 658–668, 2008.

[31] S. Kahan, T. Pavlidis, and H. S. Baird, "On the Recognition of Printed Characters of Any Font and Size," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-9, no. 2, pp. 274–288, 1987.

[32] Z. Ping and C. Lihui, "A novel feature extraction method and hybrid tree classication for handwritten numeral recognition," *Elsvier*, vol. 23, pp. 45–56, 2002.

[33] U. Garain and B. B. Chaudhuri, "Recognition of online Handwritten Mathematical Expressions," *IEEE*, vol. 34, no. 6, pp. 2366–2376, 2004.

[34] N. Arica and F. T. Yarman-vural, "An Overview of Character Recognition Focused on Off-Line Handwriting," *IEEE*, vol. 31, no. 2, pp. 216–233, 2001.

[35] N. R. Soora and P. S. Deshpande, "Review of Feature Extraction Techniques for Character Recognition," vol. 2063, no. July, 2017.

[36] T. Q. Phan, P. Shivakumara, S. Bhowmick, S. Li, C. L. Tan, and U. Pal, "Semiautomatic ground truth generation for text detection and recognition in video images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 8, pp. 1277–1287, 2014.

[37] N. R. Soora and P. S. Deshpande, "Robust Feature Extraction Technique for License Plate Characters Recognition," *IETE J. Res.*, vol. 61, no. 1, pp. 72–79, 2015.

[38] P. P. Roy, S. Roy, and U. Pal, "Multi-oriented text recognition in graphical documents using HMM," *Proc. - 11th IAPR Int. Work. Doc. Anal. Syst. DAS 2014*, pp. 136–140, 2014.

[39] M. Rani and Y. K. Meena, "An Efficient Feature Extraction Method for handwritten character recognition," *Springer-Verlag Berlin Heidelb.*, pp. 302–309, 2011.

[40] B. V. and H. B. M. Blumenstein, "A Novel Feature Extraction Technique for the Recognition of Segmented Handwritten Characters," *IEEE Comput. Soc.*, pp. 1–5, 2003.

[41] T. R. Rammohan and B. N. Chatterji, "Recognition of Distorted Kannada Characters," *IETE J. Res.*, vol. 30, no. 6, pp. 223–225, 1984.

[42] S. D. Connell and R. M. K. Sinha, "Unconstrained On-Line Devanagari Characters," *IEEE*, pp. 368–371, 2000.

[43] N. Das, A. Mollah, R. Sarkar, and S. Basu, "A comparative study of different feature sets for recognition of handwritten Arabic numerals using a Multi Layer Perceptron," *Proc. Natl. Conf. Recent Trends Intell. Comput. Kalyani*, pp. 86–92, 2006.

[44] R. Bajaj, L. Dey, and S. Chaudhury, "Devnagari numeral recognition by combining decision of multiple connectionist classifiers," *Sadhana - Acad. Proc. Eng.*

 DOI : 10.35291/2454-9150.2019.0088

Sci., vol. 27, no. PART 1, pp. 59–72, 2002.

[45] V. Bansal and R. M. K. Sinha, "Integrating knowledge sources in Devanagari text recognition system," *IEEE Trans. Syst. Man, Cybern. Part ASystems Humans.*, vol. 30, no. 4, pp. 500–505, 2000.

[46] S. Arora, D. Bhattacharjee, M. Nasipuri, D. K. Basu, and M. Kundu, "Recognition of Non-Compound Handwritten Devnagari Characters using a Combination of MLP and Minimum Edit Distance," *Int J. Comput. Sci.*, no. 4, pp. 107–120, 2010.

[47] T. Saba, A. Rehman, and G. Sulong, "Improved statistical features for cursive character recognition," *Int. J. Innov. Comput. Inf. Control*, vol. 7, no. 9, pp. 5211–5224, 2011.

[48] T. V. Ashwin and P. S. Sastry, "A font and size-independent OCR system for printed Kannada documents using support vector machines," *Sadhana - Acad. Proc. Eng. Sci.*, vol. 27, no. PART 1, pp. 35–58, 2002.

[49] A. Negi, C. Bhagvati, and B. Krishna, "An OCR system for Telugu," *IEEE*, pp. 1110–1114, 2001.

[50] B. Zhu and M. Nakagawa, "A robust method for coarse classi fi er construction from a large number of basic recognizers for on-line handwritten Chinese / Japanese character recognition," *Pattern Recognit.*, pp. 1–9, 2013.

[51] K. Cheung, S. Member, and D. Yeung, "A Bayesian Framework for Deformable Pattern Recognition With Application to Handwritten Character Recognition," *IEEE*, vol. 20, no. 12, pp. 1382–1388, 1998.

[52] R. S. R. K. and R. D. S. Samuel., "On-line Character Recognition for Handwritten Kannada Characters using Wavelet Features and Neural Classifier," *IETE J. Res.*, vol. 2063, no. March, 2015.

[53] S. Prasad, G. K. Verma, B. K. Singh, and P. Kumar, "Basic handwritten character recognition from multi-lingual image dataset using multi-resolution and multi-directional transform," *Int. J. Wavelets, Multiresolution Inf.*, vol. 10, no. 5, pp. 1–28, 2012.

[54] Y. X. and Z. S. P. Xiang, "A hybrid method for robust car plate character recognition," *IEEE Int. Conf. Syst., Man Cybern.*, pp. 4733–4737, 2004.

[55] H. Lee, S. Chen, and S. Wang, "Extraction and Recognition of License Plates of Motorcycles and Vehicles on Highways," *IEEE*, pp. 2–5, 2004.

[56] B. B. Chaudhuri, U. Pal, and M. Mitra, "Automatic recognition of printed Oriya script," *Sadhana*, vol. 27, no. February, pp. 23–34, 2002.

[57] S. Basu, N. Das, and R. Sarkar, "An MLP based Approach for Recognition of Handwritten ' Bangla ' Numerals," *Proc. 2 nd Indian Int. Conf. Artif. Intell.*, pp. 407–417, 2005.

[58] K. K. Kim, K. I. Kim, J. B. Kim, and H. J. Kim, "Learning-based approach for license plate recognition," vol. 00, no. C, pp. 614–623, 2000.

[59] J. V. Moreau, C. Olivier, Y. Lecourtier, L. Heutte, and T. Paquet, "A structural/statistical feature based vector for handwritten character recognition," *Pattern Recognit. Lett.*, vol. 19, no. 7, pp. 629–641, 1998.

[60] T. N. Yang and S. De Wang, "A rotation invariant printed Chinese character recognition system," *Pattern Recognit. Lett.*, vol. 22, no. 2, pp. 85–95, 2001.

[61] T. D. Duan, T. L. H. Du, T. V. Phuoc, and N. V. Hoang, "Building an automatic vehicle license plate recognition system," *Proc. Int. Conf. Comput. Sci. RIVF*, no. 1, pp. 59–63, 2005.

[62] S. Bag, G. Harit, and P. Bhowmick, "Recognition of Bangla compound characters using structural decomposition," *Pattern Recognit.*, vol. 47, no. 3, pp. 1187–1201, 2014.

[63] Y. Wen, J. Yan, K. M. von Deneen, P. Shi, Y. Lu, and Z. Zhou, "An Algorithm for License Plate Recognition Applied to Intelligent Transportation System," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 3, pp. 830–845, 2011.

[64] Y. P. Huang, C. H. Chen, Y. T. Chang, and F. E. Sandnes, "An intelligent strategy for checking the annual inspection status of motorcycles based on license plate recognition," *Expert Syst. Appl.*, vol. 36, no. 5, pp. 9260–9267, 2009.

[65] U. Bhattacharya and B. B. Chaudhuri, "Handwritten numeral databases of Indian scripts and multistage recognition of mixed numerals," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 3, pp. 444–457, 2009.

[66] J. Cao, M. Ahmadi, and M. Shridhar, "Recognition of handwritten numerals with multiple feature and multistage classifier," *Pattern Recognit.*, vol. 28, no. 2, pp. 153–160, 1995.

[67] B. B. Chaudhuri and U. Pal, "A complete printed Bangla OCR system," *Pattern Recognit.*, vol. 31, no. 5, pp. 531–549, 1998.

[68] G. Srikantan, S. W. Lam, and S. N. Srihari, "GRADIENT BASED CONTOUR ECODING FOR CHARACTER RECOGNITION," vol. 29, no. 7, 1996.

[69] Y. Sohn, K. Oh, and B. Kim, "A Recognition Method of the Printed Alphabet By using Nonogram Puzzle."

 DOI : 10.35291/2454-9150.2019.0088