ISSN: 0974-8571 Vol.13 No. 1 June, 2021

International Journal of Computational Intelligence in Control

Pashto Character Recognition for Low Recourse Devices in Unconstraint Environment

Muhammad Shabir[1*], Naveed Islam[1], Niamat Ullah[2], Afzal Rahman[2], Hameed Hussain[2], Kifayat Ullah[3]

[1]Islamia Collage University Peshawar, KP, Pakistan, shabiradam86@gmail.com, naveed.islam@icp.edu.pk.

[2]University of Buner, KP, Pakistan, niamatnaz@gmail.com, afzal85@uop.edu.pk, dr.hameed@ubuner.edu.pk.

[3]University of Swat, KP, Pakistan, kifayat@uswat.edu.pk

Date of Submission: 25th January 2021 Revised: 7th April 2021 Accepted: 27th June 2021

***Abstract* – Pashto languages is cursive in nature and cursive script based languages are difficult to process. The Pashto language speakers scattered and large in numbers spread around the globe. Isolated handwritten character classification and recognition techniques existed, but their performance degrade with rotated characters. The proposed work is aim to provide a generalized technique for the classification and recognition of rotated invariant handwritten characters of Pashto language. The proposed technique uses histogram of oriented gradient (HOG) features. As HOG is rotated invariant, therefore it is very useful in different types of rotation. We normalized the feature vectors to reduce time complexity. The proposed technique uses Euclidean Distance and multiple thresholds rule based classification and recognition. The proposed technique achieves maximum prediction accuracy of 93.5% for 730 characters.**

Index Terms – Pashto Characters, Rotation Invariant, HOG Features, Euclidean Distance, Multiple Thresholds Rule Based classification and Recognition.

INTRODUCTION

The character set determine the writing style of a language. The characters are symbols, which are combines to form a group. The words of languages are composed of these characters. Most of the languages use the Arabic, Roman and Chines characters sets while some languages use the modified version of these character sets. Combination of characters form different words of the language if exists in a standard dictionary. The characters are connecting at difference position with each other's to for a word i.e. Pashto, Urdu, Dari etc. The connection of characters are called joints, found at the front, back or from both side. The concatenation of characters is also used to form a words i.e. English language. The concatenation is takes place in a sequence without make joint with next or previous character.

Pashto characters connect vertically with each other at different position. Direction is another important property found in the writing script of a language. The direction of writing Pashto language script is from left to right and from top to bottom. The direction and connecting property produces a lot of curves and gradients. Characters of Pashto language consist of one or more strokes. The first stroke is called Primary stroke or some time called the skeleton of the character. Most of the characters has more than one stroke, in which first stroke is called the primary and all other stroke are called secondary. A stroke is the continuous on pixels in a trajectory from first pixel to last pixel. The primary or skeleton of a character is the longest stroke in which the numbers of on pixels are large in number as compare to other continuous on pixels.

Most of the cursive script characters recognition techniques use machine learning algorithms (MLA). Machine learning algorithms has two broad category i.e. Supervised and unsupervised. In unsupervised MLA data is directly feed without labeling. Unsupervised MLA's make groups of similar data. In supervised MLA's data are labeled explicitly and then feed into algorithm for classification. Supervised MLA's are heavily use for character classification and recognition. Machine learning algorithms are computationally expensive and not efficient without normalization.

The proposed technique can work for all cursive script based languages but hereby it is tested on Pashto characters.

Most of the existing systems are imposed constraint on user for writing input characters. The system is not providing the same or nearest recognition accuracy if the input characters are slightly rotated. In proposed technique the recognition of characters, take place in a control environment, which is very less affected with different types of rotation.

RELATED WORK

Handwritten character recognition systems has focused for the last two to three decades. Large number of techniques are developed for single language characters recognition [1]. Pashto language has less reported work in this regard.

Existing techniques have also limitations in terms of computational speed, writing constraints on user and memory occupied. `

A weighted linear classifier with local feature is used by [2] for online handwriting Urdu isolated characters recognition. In this method weights are assigned to each feature vector and uses assumption which is not a good idea. It is a single language characters recognition method and claim 92.8% accuracy. This technique imposed constraints on user while writing.

Fuzzy rule based technique is used for character recognition by [3]-[5] to classify online Urdu characters. It is very hard to define too many Fuzzy rules explicitly for each character. These techniques are also bound to a single language and depend on baseline.

Hidden Markov model based method is heavily used for handwriting character recognition. In this method models of each character is prepared and then stored in a dataset for classification. It is a statistical process in which probability is used in each stat for transition to the next stat. Due to probability this method is very closed to dynamic Bayesian network. HMM based techniques for character recognition are not rotated, scaled and shift invariant and never used for multi-language characters recognition [6]-[15].

Template matching technique is heavily used in different other area i.e. robotics and image processing. In template matching technique templates for each character is prepared. The templates are then used in recognition phase for unseen character [16]-[21]. These template matching based techniques only support single language characters, which are not geometric invariant.

Support vector machine (SVM) based classification and recognition has been found in many research papers. It is a supervise machine learning technique used different kernel tricks. The different form of SVM is uses for different problem depend on the type of problem. The SVM technique can also be used for regression. The linear kernel based SVM is used for the separation of two different classes with the help of hyperplane. The hyperplane is generated between two classes with the help of margins. Margin in hyperplane are distances between hypothesis line and margin vectors. Other SVM kernels uses the same margin concepts and separate data in higher dimensions in a non-linear way. These kernels i.e. Polynomial kernel, Hybrid kernel, Sigmoid kernel, Exponential kernel and Redial basis kernel are efficient for complex boundary separation. SVM in different form has used for characters classification and recognition [22]-[25]. All these SVM based method has not tolerated for rotation in variation.

Artificial neural network (ANN) is a supervise machine learning algorithm inspired from human nervous system. Different variation of ANN is used to classify characters on the basis of features. Different forms of ANN are feed forward neural network, feed backward neural network, recurrent neural network and convolution neural network etc. In Artificial neural network weights are multiplied with features data and then perform computation by using an activation function. For character recognition [26] uses ascender and descender lines to extract features and ANN to classify characters.

Genetic algorithm inspired from human genetic system and performs mutation and permutation to find global maxima. In this method fattest candidate survived for next generation. This method performs different computation steps on characters related data [27].

Naïve Bayes is a supervise machine learning technique, which is based on probability. This technique has used for character classification [28]. Features of handwriting characters are feed into Naïve Bayes classifier for classification. Naïve Bayes use interior and posterior probability of character features. Naïve Bayes based classification is simple but very effective and fast.

We proposed a rotated invariant handwriting characters technique for Pashto character classification and recognition. The consideration of rotation invariant property is very important [29] as humans' produces different amount of rotation in handwriting .The proposed system do not imposed constraint on handwriting rotation variation, ascender and descender line.

PROPOSED METHOD

User interaction with smart devices are very easy. Every common person can easily use the smart devices very conveniently without any hesitation. These devices are easy to carry from place to place along with functionality. Millions of applications have been developed to facilitate end users. A common user understand and communicate in native languages and the symbols related to that language. There are large number of peoples that communicate only in regional language. They do not understand English and other languages except their regional language. This problem is very common in Pashtun community. To take the advantage of touch sensitive devices, we proposed a technique for Pashto handwritten characters classification and recognition. The proposed technique recognizes rotated invariant handwritten characters of Pashto languages. This technique consists of Smoothing, Features extraction, weighted normalized vector and Recognition as shown in Figure 1.

In proposed technique, Pashto handwritten characters are taken and preprocess it to remove the salt and pepper noise. HOG features vectors are extracted which are invariant to the rotation. Eigen values and Eigen vectors are used to normalize HOG feature vectors to boost the computational speed. In proposed technique, Euclidean distance is computed for each features vector, which describe a character. In testing phase unseen character stroke is taken and performs computation steps similar to training phase. Euclidean distance from training and testing phases are matched for decision. In recognition phase rules are define to decide a label for unseen data.

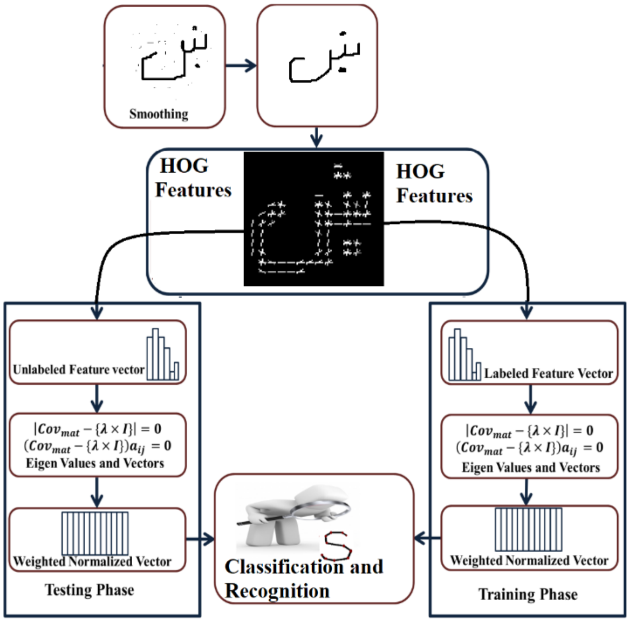


FIGURE 1: PROPOSED TECHNIQUE ARCHITECTURE

Handwriting characters contain salt and pepper noise. The noise are affect the recognition accuracy and need to reduce. We reduce the salt and paper noise by using median filter. Median filter is a statistical method to reduce the salt and pepper noise. In this filter, the median value is find and replace in the pixel below the filter. Median filter shows very good results as shown in Figure 2.

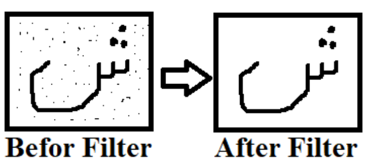


FIGURE 2: REDUCTION OF NOISE

FEATURES EXTRACTION

In the proposed Technique, we select Histogram of Oriented Gradient (HOG) as feature to extract data. Features extracting from Pashto handwritten characters play a vital role in classification and recognition. Selecting affective features from a set of features is very important [30]. We choose to extract Histogram of Gradient (HOG) for each handwritten character. The reason of choosing histogram of gradient is the rotation invariant nature. It help in the classification and recognition of geometrical invariant characters. The initial size of an image is 128×128 as shown in Figure 3.

FIGURE 3: PASHTO CHARACTER OF SIZE 128×128

We resizes the character images to 64×64 as shown in Figure 4. Resize the images reduce the computational complexity.

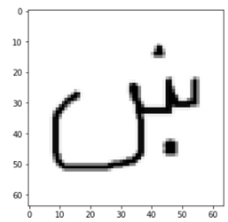


FIGURE 4: RESIZE TO 64×64

We set the size of HOG cell into 8×8 because we do not want to miss detail. The number of cell in each block is configured are 2×2. In the proposed system, the number of orientation histogram bins are nine. We use logical scaler value for orientation, which create evenly spaced in bins between 0 and 180 degrees. The HOG image for a Pashto handwritten character is shown in Figure 5.

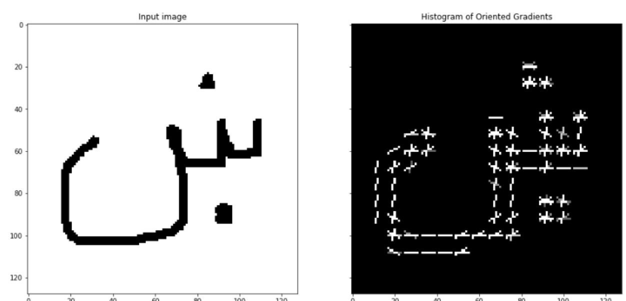


FIGURE 5: INPUT IMAGE AND THE RESPECTIVE HOG IMAGE.

By applying HOG method for feature extraction, two vectors are constructed that is gradient magnitude and gradient direction. The values in these two vectors are find out by using the pixel values in X direction and in Y direction. The X direction value is calculated by using (1) and the Y direction values is computed by using (2).

$$X_D = |Pl - Pr| \qquad (1)$$
$$Y_D = |Pu - Plw| \qquad (2)$$

In (1) X_D is the value in X direction, *Pl* is the value of left pixel and *Pr* is the value of right pixel. Y_D in (2) is the value in Y direction, *Pu* is the value of upper pixel and *Plw* is the value of right pixel.

The vector for gradient magnitude is computed by using (3) and the vector gradient direction is computed by using (4).

$$G_m = \sqrt{X_D^2 + Y_D^2} \qquad (3)$$

$$G_d = \frac{X_D}{Y_D} \qquad (4)$$

In (3) Gm is the gradient magnitude value for gradient magnitude vector, $X_D^2$is the square of X direction value and $Y_D^2$is the square of Y direction value. In (4) G_d is the gradient direction value for gradient direction vector, X_D is the value of X direction and Y_D is the value of Y direction. The total HOG blocks in our case is 225 i.e. (15×15) and each block has 16×16 HOG cell. It means that each block has 36×1 vectors. The total size of our feature vector $\vec{V}$ is 225×36 = 8100. We normalized the HOG feature vectors by using the following method.

In ordered to make the proposed technique more efficient on low resource devices, the features vector $\vec{V}$ belong to a character is normalized. Eigen values are computed and then used to produce weighted normalized vector. The whole process demonstrated steps wise in Figure 6.

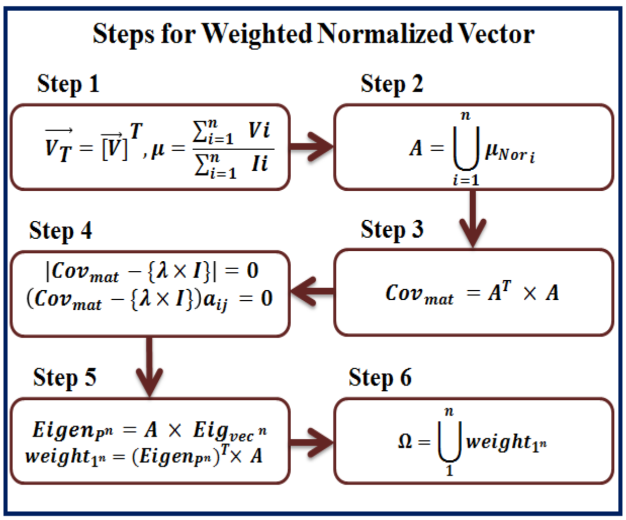


FIGURE 6: STEP WISE WEIGHTED NORMALIZED VECTOR COMPUTATION

Step 1: Each character features vector is transposed from 1×N to N×1 as follows in (5)

$$\overrightarrow{V\ T} = [\vec{V}]T \qquad (5)$$

For each $\overrightarrow{V\ T}$ mean value "μ" is computed as follows.

$$\mu = \frac{\sum_{i=1}^{n} \quad Vi}{\sum_{i=1}^{n} \quad Ii} \qquad (6)$$

Step 2: Mean normalized vector μ_{Nor} for each feature vector $\vec{V}$ is computed as follows.

$$\mu_{Nor} = \overrightarrow{V\ T} - \mu \qquad (7)$$

After computing all characters normalized vector total normalized matrix A is constructed by concatenating μ_{Nor} as given below in (8).

$$A = \cup_{i=1}^{n} \mu_{Nor_i} \qquad (8)$$

Step 3: Total covariance matrix is computed with the help of (9).

$$Cov_{mat} = A \qquad (9)$$

Covariance matrix is an arrangement of rows and columns of covariant values. Mathematically a covariance value can only be measured between two features vector as shown in (10).

$$Cov_{val} = 1/n \sum_{i=1}^{n} \quad (x - \mu_x)\,(y - \mu_y)^T \qquad (10)$$

$\overrightarrow{V\ x}$ and $\overrightarrow{V\ y}$ are features vector, n is the total number of data points in a feature vector, μ_x is the mean of $\overrightarrow{V\ x}$ feature vector, μ_y is the mean of $\overrightarrow{V\ y}$ feature vector and T indicates the transpose operation. If a dataset consists of two features vector $\overrightarrow{V\ x}$ and $\overrightarrow{V\ y}$ then all possible covariance values will be $Cov_{val}(\overrightarrow{V\ x}, \overrightarrow{V\ x})$, $Cov_{val}(\overrightarrow{V\ x}, \overrightarrow{V\ y})$, $Cov_{val}(\overrightarrow{V\ y}, \overrightarrow{V\ x})$ and $Cov_{val}(\overrightarrow{V\ y}, \overrightarrow{V\ y})$. Covariance of a feature vector with itself becomes variance. Covariance matrix for n features vector will be n×n i.e.

$$Cov_{mat} = [\,Cov_{val}(\overrightarrow{V\ x}, \overrightarrow{V\ x})\ Cov_{val}(\overrightarrow{V\ x}, \overrightarrow{V\ y}), \ldots\ Cov_{val}(\overrightarrow{V\ x}, \overrightarrow{V\ n}),$$
$$Cov_{val}(\overrightarrow{V\ y}, \overrightarrow{V\ x})\ Cov_{val}(\overrightarrow{V\ y}, \overrightarrow{V\ y})\ \ldots Cov_{val}(\overrightarrow{V\ y}, \overrightarrow{V\ n})\ \ldots\ \ldots\ \ldots$$

.

Step 4: Eigen values are computed from covariance matrix as follows.

$$|Cov_{mat} - \{\lambda \times I\}| = 0 \qquad (11)$$

In (11) Cov_{mat} is covariance matrix of features dataset, I is identity matrix, λ is Eigen value to be calculated and || is determinant. The number of rows and columns in identity matrix should be equal to the number of rows and columns in covariance matrix. If Cov_{mat} is n × n, then total number of λ will also be n i.e. λ_1, λ_2, ….. λ_n. Eigen vector is calculated by using (12).

$$(Cov_{mat} - \{\lambda \times I\})a_{ij} = 0 \qquad (12)$$

In (12) Cov_{mat} represents covariance matrix, λ represents one of the Eigen value, I indicates identity matrix and a_{ij} represent features vector i.e. a_{11}, a_{12}, a_{21}, a_{22} ….. a_{nn}. Equation (12) gives n system of equations if Cov_{mat} has n number of rows or columns. By applying rule of linear algebra on (12), Eigen vectors are computed i.e.

$Eig_{vec} = [a_{11}\ a_{12}\ \cdots\ a_{1n}\ a_{21}\ a_{22}\ \ldots\ a_{2n}\ \vdots\ \vdots\ \ddots\ \vdots\ a_{n1}\ a_{n2}\ \cdots\ a_{nn}\ ]$.

Step 5: Normalized projection matrixes $Eigen_{p^n}$ is computing for vector A by using (13).

$$Eigen_{p^n} = A \times Eig_{vec^n} \quad where\ n = 1,2,3 \ldots \quad (13)$$

Weighted normalized vectors $weight_{1^n}$ for character are computed by using (14).

$$weight_{1^n} = \times\ A \quad where\ n = 1,2,3 \ldots \quad (14)$$

Step 6: Weighted normalized vectors $weight_{1^n}$ are concatenated to find a single weighted normalized vector Ω for each character as shown in (15).

$$\Omega = \cup_1^n weight_{1^n} \quad where\ n = 1,2, \ldots \quad (15)$$

Classification and Recognition

In classification and recognition phases, both has calculate weighted normalized vector and Euclidean distance. Multiple threshold values are set to recognize a character as shown in Figure 7. The four steps below are apply on feature vectors for both classification and recognition phase. The only difference is the labeling of feature vectors. We explain these four steps with respect to recognition.

Step 1: Features related to a character trajectory are taken called $\vec{V}$ without being explicitly labeled and transpose it for further computation as shown in (16).

$$Test_T = \quad (16)$$

Step 2: Mean normalized vector is computed for $Train_T$ through (17).

$$Test_{Nor} = \quad (17)$$

Step 3: Weighted normalized vectors $weight_{T^n}$ for test data are computed with the help of (18).

$$weight_{T^n} = \times\ Train_{Nor} \quad (18)$$
$$where\ n = 1,2,3 \ldots$$

Step 4: Weighted normalized vectors for testing data $weight_{T^n}$ are concatenated in a single weighted vector called Ω_{Test} as shown in (19)

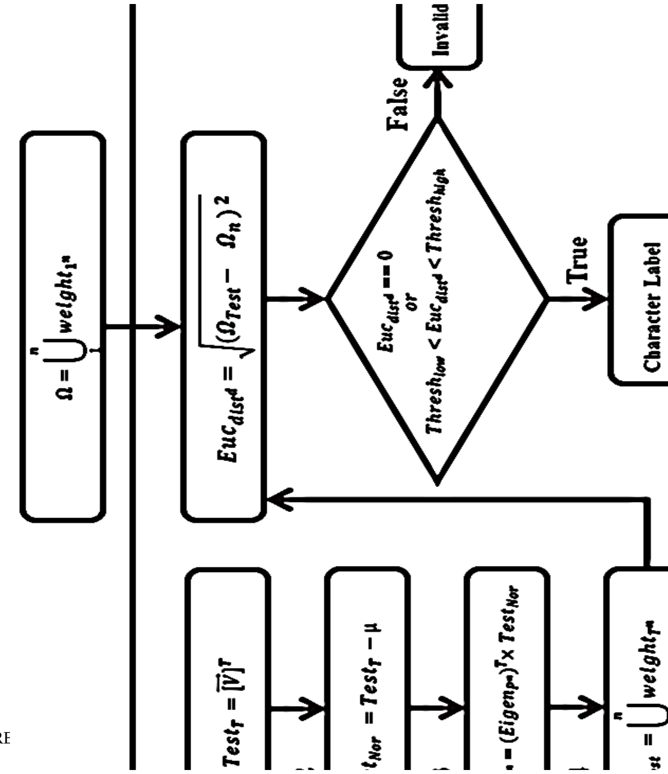
$$\Omega_{Test} = \cup_1^n weight_{T^n} \quad where\ n = 1,2,3 \ldots \quad (19)$$

To recognized unlabeled characters data, Euclidean distance is computed against every Ω as shown in (20).

$$Euc_{dist^d} = \sqrt{(\Omega_{Test} - \Omega_n)^2} \quad (20)$$

Where n is set of characters and d = {1, 2, 3…}. After computing Euclidean distance, rule are define, which based on multiple threshold values. If one of the Euclidean is distance, value Euc_{dist^d} satisfy the condition that will be the recognized character as shown in (21).

$$Return = \{Euc_{dist^d}, if\ Euc_{dist^d} = = 0\ or\ Thresh_{low} < Euc_{dist^d} < Thresh_{high} \quad (21)\ Invalid\ Character,\ otherwise\ where\ d = features\ vectors$$



FIGURE

Result and Discussions

To recognize rotated invariant Pashto handwriting characters, data are collected from teachers, students and peons. 730 different samples of handwritten invariant Pashto characters are taken. The proposed technique gives 93.5% accuracy for teacher handwriting characters, 91.9% for students and 88.6% for peons as shown in table 1. The recognition accuracy is determine by using (22).

$$Accuracy\ in\ \% = \frac{True\ Recognized\ Sample}{Total\ Number\ of\ Sample} \times 100 \quad (22)$$

TABLE 1: RECOGNITION ACCURACY OF PROPOSED SYSTEM

Users	Education	Total Samples	Correctly Recognized Samples	Accuracy in %
Teachers	**Post Graduated**	**730**	**683**	**93.5%**
Students	**10th Standard**	**730**	**679**	**91.9%**

Peons	Primary School	730	647	88.6%

The variation in recognition accuracy is due to the education level. In this case, teachers has better writing than students and students have better writing than peons. Most of the peons have eight standers of education. They write some characters which difficult for humans to understand.

Conclusion

As there are many techniques proposed for handwriting character recognition but most of them are based on ascenders, descender and base lines. Most of the existing methods are for roman script based language. The proposed technique has no constraint on user while writing. It recognized rotated characters for Pashto language. The proposed method is very fast and elegant due weighted normalized vectors. Existing classifiers i.e. SVM and ANN etc are computationally expensive for low recourse devices. The proposed method uses Euclidean distance and rules for recognition. The proposed technique achieved remarkable results for different types of user even for less experience people. The main advantage is the feasibility of the proposed system for low resource devices.

References

[1] Puri, S., & Singh, S. P. (2016, January). Text recognition in bilingual machine printed image documents—Challenges and survey: A review on principal and crucial concerns of text extraction in bilingual printed images. In 2016 10th International Conference on Intelligent Systems and Control (ISCO) (pp. 1-8). IEEE..

[2] Shahzad, N., Paulson, B., & Hammond, T. (2009, February). Urdu Qaeda: recognition system for isolated urdu characters. In Proceedings of the IUI Workshop on Sketch Recognition, Sanibel Island, Florida.

[3] Malik, S., & Khan, S. A. (2005, September). Urdu online handwriting recognition. In Proceedings of the IEEE Symposium on Emerging Technologies, 2005. (pp. 27-31). IEEE.

[4] Razzak, M. I., Anwar, F., Husain, S. A., Belaid, A., & Sher, M. (2010). HMM and fuzzy logic: a hybrid approach for online Urdu script-based languages' character recognition. Knowledge-Based Systems, 23(8), 914-923.

[5] Liwicki, M., Graves, A., Fernàndez, S., Bunke, H., & Schmidhuber, J. (2007). A novel approach to on-line handwriting recognition based on bidirectional long short-term memory networks. In Proceedings of the 9th International Conference on Document Analysis and Recognition, ICDAR 2007.

[6] Toselli, A. H., Pastor, M., & Vidal, E. (2007, June). On-line handwriting recognition system for tamil handwritten characters. In Iberian Conference on Pattern Recognition and Image Analysis (pp. 370-377). Springer, Berlin, Heidelberg.

[7] Biadsy, F., El-Sana, J., & Habash, N. Y. (2006). Online arabic handwriting recognition using hidden markov models.

[8] Liwicki, M., & Bunke, H. (2007, September). Combining on-line and off-line systems for handwriting recognition. In Ninth International Conference on Document Analysis and Recognition (ICDAR 2007) (Vol. 1, pp. 372-376). IEEE.

[9] Connell, S. D., & Jain, A. K. (2002). Writer adaptation for online handwriting recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(3), 329-346.

[10] Patil, P., & Ansari, S. (2014). Online handwritten devnagari word recognition using HMM based technique. International Journal of Computer Applications, 95(17).

[11] Nakai, M., Akira, N., Shimodaira, H., & Sagayama, S. (2001, September). Substroke approach to HMM-based on-line Kanji handwriting recognition. In Proceedings of Sixth International Conference on Document Analysis and Recognition (pp. 491-495). IEEE.

[12] Pechwitz, M., & Maergner, V. (2003, August). HMM based approach for handwritten Arabic word recognition using the IFN/ENIT-database. In Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings. (Vol. 3, pp. 890-890). IEEE Computer Society.

[13] Artieres, T., Marukatat, S., & Gallinari, P. (2007). Online handwritten shape recognition using segmental hidden markov models. IEEE transactions on pattern analysis and machine intelligence, 29(2), 205-217.

[14] Amor, N. B., & Amara, N. E. B. (2006). Multifont Arabic Characters Recognition Using HoughTransform and HMM/ANN Classification. J. Multim., 1(2), 50-54.

[15] Sternby, J., Morwing, J., Andersson, J., & Friberg, C. (2009). On-line Arabic handwriting recognition with templates. Pattern recognition, 42(12), 3278-3286.

[16] Aparna, K. H., Subramanian, V., Kasirajan, M., Prakash, G. V., Chakravarthy, V. S., & Madhvanath, S. (2004, October). Online handwriting recognition for Tamil. In Ninth international workshop on frontiers in handwriting recognition (pp. 438-443). IEEE.

[17] Santosh, K. C., & Nattee, C. (2009). A comprehensive survey on on-line handwriting recognition technology and its real application to the Nepalese natural handwriting. Kathmandu University Journal of Science, Engineering, and Technology, 5(I), 31-55.

[18] Garain, U., Chaudhuri, B. B., & Pal, T. T. (2002, August). Online handwritten Indian script recognition: a human motor function based framework. In Object recognition supported by user interaction for service robots (Vol. 3, pp. 164-167). IEEE.

[19] Tanaka, H., Nakajima, K., Ishigaki, K., Akiyama, K., & Nakagawa, M. (1999, September). Hybrid pen-input character recognition system based on integration of online-offline recognition. In Proceedings of the Fifth International Conference on Document Analysis and Recognition. ICDAR'99 (Cat. No. PR00318) (pp. 209-212). IEEE.

[20] Connell, S. D., & Jain, A. K. (2001). Template-based online character recognition. Pattern Recognition, 34(1), 1-14.

[21] Izadi, S., & Suen, C. Y. (2008, December). Online writer-independent character recognition using a novel relational context representation. In 2008 Seventh International Conference on Machine Learning and Applications (pp. 867-870). IEEE.

[22] Swethalakshmi, H., Jayaraman, A., Chakravarthy, V. S., & Sekhar, C. C. (2006, October). Online handwritten character recognition of Devanagari and Telugu Characters using support vector machines. In Tenth international workshop on Frontiers in handwriting recognition. Suvisoft.

[23] Kumar, D., & Rana, N. (2011). Speech Synthesis System for Online Handwritten Punjabi Word: An Implementation of SVM & Concatenative TTS. International Journal of Computer Applications, 975, 8887.

[24] Nakkach, H., Hichri, S., Haboubi, S., & Amiri, H. (2016, March). Hybrid approach to features extraction for online arabic character recognition. In 2016 13th International Conference on Computer Graphics, Imaging and Visualization (CGiV) (pp. 253-258). IEEE.

[25] Jaeger, S., Manke, S., Reichert, J., & Waibel, A. (2001). Online handwriting recognition: the NPen++ recognizer. International Journal on Document Analysis and Recognition, 3(3), 169-180.

[26] Potrus, M. Y., Ngah, U. K., & Ahmed, B. S. (2014). An evolutionary harmony search algorithm with dominant point detection for recognition-based segmentation of online Arabic text recognition. Ain Shams Engineering Journal, 5(4), 1129-1139.

[27] Abandah, G. A., Younis, K. S., & Khedher, M. Z. (2008, February). Handwritten Arabic character recognition using multiple classifiers based on letter form. In Proceedings of the 5th International Conference on Signal Processing, Pattern Recognition, and Applications (SPPRA) (pp. 128-133).
[28] Amalia, A., Sharif, A., Haisar, F., Gunawan, D., & Nasution, B. B. (2018, October). Meme opinion categorization by using optical character recognition (OCR) and naïve bayes algorithm. In 2018 Third International Conference on Informatics and Computing (ICIC) (pp. 1-5). IEEE.
[29] Jan, Z., Shabir, M., Khan, M. A., Ali, A., & Muzammal, M. (2016). Online Urdu handwriting recognition system using geometric invariant features. Nucleus, 53(2), 89-98.
[30] Jan, Z., Shabir, M., KHAN, M., SHAH, S., & Baloch, B. (2015). Features used in online handwriting and signature recognition systems: A survey. Sindh University Research Journal-SURJ (Science Series), 47(4).