Robotic Handwritten Kannada Character Recognition using Neural Network

Shakunthala B. S, C S Pillai

Abstract*: Data preparing and the board is basic now a days. In this paper, programmed preparing of structures written in Kannada language is considered. A reasonable pre-preparing procedure is introduced for separating written by hand characters. Essential Component Analysis (PCA) and Histogram of arranged Gradients (HoG) are utilized for highlight extraction. These highlights are sustained to multilayer feed forward back spread neural system for arrangement. Just 57 characters are utilized for acknowledgment. Exhibitions of two highlights are looked at for changed number of classes. Hoard is found to have preferred acknowledgment exactness over PCA as number of classes expanded. This is actualized in Visual Studio 2010 utilizing Open CV library.

Keywords : Back Propagation Neural Network, Form Processing, Histogram of Gradients, Kannada Script, Principal Component Analysis.

I. INTRODUCTION

The robotization of transcribed structure preparing is drawing in concentrated research enthusiasm because of its wide application and decrease of manual work. In Indian setting, numerous associations will gather the information on paper based structures. Programmed handling of these structures is a procedure of catching the data put away in the structures and changing over it into electronic (machine coherent) group. Written by hand character acknowledgment massively to the progression of robotization process. This is grouped into disconnected and online acknowledgment. For Automatic Form Processing (AFP) disconnected acknowledgment technique is utilized since it includes programmed change of content in a picture into letter codes which are usable inside PC and content handling applications [1]. AFP incorporates total output of a structure utilizing scanner. The filtered picture at that point experiences different pre-handling activities, character division and acknowledgment of written by hand characters. India is a multi-lingual and multi-content nation containing eighteen authority dialects, Kannada is one among them. A few works has been improved the situation the acknowledgment of written by hand Kannada characters. The major pre preparing ventures of AFP incorporates edge identification, morphological tasks to make it reasonable for division. Division isolates the picture content archives into lines, words and the characters. Thungamani M and RamakanthKumar P [2] talk about two division strategies, for example, established methodology and all-encompassing methodology. In established methodology, the info picture is fragmented into sub pictures. In all encompassing methodology, the characters are perceived without analyzation. Mamatha H.R and Srikanatamurthy K [3], proposed a division conspire utilizing projection profiles. Morphological activities are utilized to evacuate the commotion. After this content lines are removed utilizing flat projection profile, words and characters are separated utilizing vertical projection profiles. India is a multi-lingual and multi-content nation involving eighte authority dialects, Kannada is one among them. A few works has been improved the situation the acknowledgment of written by hand Kannada characters. The major pre handling ventures of AFP incorporates edge location, morphological tasks to make it appropriate for division. Division isolates the picture content records into lines, words and the characters. Thungamani M and RamakanthKumar P [2] talk about two division procedures, for example, traditional methodology and all encompassing methodology. In established methodology, the information picture is portioned into sub pictures. In comprehensive methodology, the characters are perceived without analyzation. Mamatha H.R and Srikanatamurthy K [3], proposed a division plot utilizing projection profiles. Morphological activities are utilized to evacuate the commotion. After this content lines are removed utilizing level projection profile, words and characters.

There are different sorts of division plans, out of which the morphological tasks and certain commotions are to be expelled which character perceiving through picture division process, since the characters are composed writings with letters in order, words, sentences and lines of the content. So regularly these things are portioned independently to make it plausible to the character to be perceived. Hence it is important to make the characters to make it conspicuous which is done through picture division. Consequently after the writings are being extricated it must be anticipated. This is finished utilizing neural system ideas.

Kannada content has substantial number of character set. This may decrease the acknowledgment precision and increment the computational expense. To maintain a strategic distance from this issue a calculation has been proposed to diminish image set [4], where the vowel modifiers (kagunitha) and consonant modifiers (vattakshara) which are not associated with base characters are considered as discrete classes. Devanagari content has comparative attributes as Kannada content like vowel modifiers, consonant conjuncts and so forth,. The acknowledgment of this content comprises of three stages: division, disintegration, i.e,. breaking down a composite character into base part and modifier parts; and acknowledgment [5]. Just little subsets of compound characters (upper and lower signs) are considered for the acknowledgment. Numerous Arabic letters additionally share normal essential shapes, which contrasts just in the quantity of dabs and the dabs or above or

Revised Manuscript Received on August 05, 2019.

Shakunthala B S*, Assistant Professor, Department of Information Science & Engineering, Kalpataru Institute of Technology, Tiptur, Karnataka, India.

Dr. C S Pillai, Professor, Department of Computer Science & Engineering, ACS college of engineering, Bangalore, Karnataka, India.

beneath the essential shape. A study on disconnected acknowledgment of Arabic penmanship acknowledgment is introduced in [6]. Diverse division, include extraction and acknowledgment motors utilized for OCR are additionally talked about. A review of character acknowledgment strategies as for the disconnected character acknowledgment frameworks, for example, pre-handling, division, portrayal, acknowledgment and post-preparing techniques are displayed in [7]. Shape based highlights, for example, Fourier descriptors what's more, chain codes are utilized for the acknowledgment of transcribed Kannada characters (vowels and numerals) are talked about in [8]. Bolster Vector Machine (SVM) is utilized for acknowledgment reason and a precision of 95% is gotten. A short overview on disconnected acknowledgment of Devanagari content is introduced in [9]. Execution of various component extraction strategies utilizing distinctive classifiers is organized. Angle and PCA based highlights with PCA, SVM and Neural Network classifiers are found to have better acknowledgment precision. Advancement of two databases for two famous Indian contents Devanagari and Bangla for numeral acknowledgment is exhibited in [10]. This uses a multistage course acknowledgment conspire utilizing wavelet-based multi-goals portrayals and multi-layer perceptron (MLP) classifiers to accomplish higher recogniti precision. This is then used to the acknowledgment of blended numerals for three Indian contents, for example, Devanagari, Bangla and English. Unconstrained written by hand acknowledgment of Kannada characters utilizing vast dataset of 200 examples utilizing ridgelet change is examined in [11]. To diminish the component of highlight vector PCA is utilized. It is discovered that ridgelet highlights offered promising outcome than PCA. A zone based strategy for the acknowledgment of manually written Kannada vowels and consonants is displayed in [12]. Character picture is partitioned into 64 non-covered zones and from each zone decipher codes are registered. SVM is utilized as classifier and an exactness of 87.24% is accomplished. Writing records few papers on Kannada character acknowledgment. Selection of techniques for highlight extraction is vital for accomplishing proficient character acknowledgment for vast classes. In this paper, Kannada penmanship acknowledgment for programmed structure preparing is considered. PCA and HoG are utilized for highlight extraction. Exhibitions of highlights are analyzed for 57 classes.

II. KANNADA SCRIPT

Kannada letters in order were produced from th descendents of Brahmi content, for example, Kadamba and Chalukya contents. Manually written character acknowledgment of Kannada characters is a testing errand in light of its extensive dataset, shape likeness among characters a non-uniqueness in the portrayal of diacritics. Kannada content has 49 essential characters: 15 vowels and 34 consonants as appeared in Fig. 1(a) and Fig. 1(b). Every one of the vowels, change essential consonants to frame a compound character or kagunita as appeared in Fig. 1(c). Furthermore a consonants accentuation glyph called consonant conjuncts appeared in Fig. 1(d) [vattakshara], exists for every one of the 34 consonant. In this way, all out number of conceivable blends is as appeared TABLE

C D E F G H IÄ J K L M N O CA CB
(a) Vowels

PÀ R UÀ WÀ Y
ZÀ bÀ d gÀhÄ k
l oÀ qÀ qsÀ t
vÀ xÀ zÀ zsÀ £À
¥À ¥sÀ § ¨sÀ ªÀÄ
AiÀÄ gÀ ® ªÀ ±À µA ¸À ºÀ ¼À

(b)Consonants

PÀ PÁ Q QÃ PÀÄ PÀÆ PÀÈ PÉ PÉÃ PÉÊ PÉÆ PÉÆÃ
PË PÀA PÀB
(c)Vowel modified consonants

Ì Í Î Ï Ð Ñ Ò Ó Ô Õ Ö × Ø Ù Ú Û Ü Ý Þ ß à á â ã ä å ç è é
ê ë ì í î
(d)Consonant conjuncts

Fig. 1: Basic Kannada character set

III. LITERATURE SURVEY

[11] Radial Basis Function (RBF) neural Network has been executed on eight directional estimations of angle highlights for manually written Hindi character acknowledgment. The character acknowledgment framework was prepared by utilizing diverse examples in various penmanship styles gathered of different individuals of various age gatherings. The Radial Basis Function coordinate with one info and one yield layer has been utilized for the preparation of RBF Network. Trial has been performed to contemplate the acknowledgment exactness, preparing time and characterization time of RBF neural system. The acknowledgment precision, preparing time and characterization time accomplished by actualizing the RBF organize have been contrasted and the outcome accomplished in past related work for example Back engendering Neural Network. Similar outcome demonstrates that the RBF with directional component gives marginally less acknowledgment exactness, decreased preparing and characterization time.

[12] Character debasement is a noteworthy issue in character acknowledgment. The greater part of the recorded archives are corrupted because of outward factors like obscuring, skew, foundation clamor and so forth and characteristic variables like contortion, broken characters, contacting characters and so on. In this paper we propose a novel methodology of modifying the broken characters and after that utilizing neural system for acknowledgment. Kannada characters are hard to perceive because of their confounded shapes and braking of characters makes it considerably more troublesome. The broken characters are in this way revamped utilizing end guide calculation toward evacuate brokenness and a solitary layer neural system is utilized for grouping. An acknowledgment exactness of 98.9% was accomplished for broken characters on artificially produced informational collections.

[13] Hindi is the normal and most well known language in the nations, for example, India, Nepal and so forth. Individuals utilize this language for discussion as well as in their vehicles tags, archives, sign sheets, manually written notes

and so on. As of late, numerous methodologies have been proposed for Hindi character acknowledgment and different applications, for example, content to discourse interpreter, programmed tag acknowledgment and so forth are proposed for these. Some computationally costly methodologies have accomplished attractive precision however for light processing gadgets, acknowledgment of written by hand characters is as yet difficult assignment. This paper proposes a methodology for acknowledgment of written by hand Devanagari character acknowledgment. The shape change of the character in Devanagari content is displayed by variation of bends. These characters are recognized utilizing highlight extraction in piecewise way. The picture parceling strategy is utilized for piecewise histogram of situated inclinations (HOG) highlights extraction. To prepare the neural system, a component vector contain HOG highlights of all segments is utilized. The proposed methodology accomplishes the limit of 99.27% order precision in preparing and can perceive the distinctive transcribed Devanagari characters with a normal exactness of 97.06%. The proposed methodology might be helpful in the application for visually impaired individuals to peruse the manually written substance.

[14] Devanagari content has character set with rich auxiliary highlights that makes the acknowledgment of unconstrained written by hand Devanagari characters troublesome However, these highlights can be utilized to partition the characters into various classes. This paper presents couple of strategies for improving the acknowledgment precision at pre-arrangement organize, highlight extraction stage and acknowledgment organize. At first, the pre-characterization of the characters is done into various classes utilizing different auxiliary highlights. At that point highlights are separated utilizing improved component extraction strategies. At last, the acknowledgment is finished utilizing neural system. In this paper, distinctive neural systems are actualized and their exhibitions are dissected.

[15] Contemplate intends to explore the execution of convolutional neural systems (CNN) on PHCR issues. To examine the execution of CNN techniques, a dataset of Persian manually written characters has been utilized as ground truth information. The dataset components changed over into pictures with the extent of 64 × 64 picture components (pixels). To elucidate the outperformance of proposed strategy, it is contrasted and prominent traditional methods in PHCR issues. The outcomes demonstrate that the execution of traditional techniques isn't great on PHCR. Notwithstanding ordinary techniques, two sorts of CNN strategies have been actualized. Single convolutional neural system (SCNN) has been actualized dependent on the straightforward structure of CNN (LeNet-5). In addition, the stowing worldview connected on CNN to broaden it into outfit convolutional neural system (ECNN) with an assortment of system parameters. The outcomes demonstrate that, in spite of the fact that the ECNN beat SCNN (accuracy=97.1%) in exactness, SCNN can perceive Persian transcribed characters with tolerable multifaceted nature and reasonable precision (accuracy=96.3%).

IV. FORM PROCESSING METHODOLOGY

Programmed Form Processing framework includes Image procurement utilizing scanner, pre-handling of examined structure, just written by hand character extraction, transcribed character division, character acknowledgment and capacity as appeared in the layout of the Birth declaration is made with all the required information fields. The candidate is then taught to fill the structure in Kannada language with all the base characters in the upper box and conjuncts in the lower box. The filled structure is then checked utilizing flatbed examining.

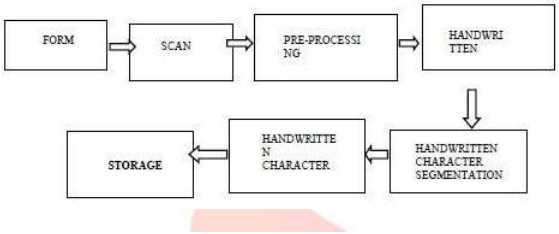

Fig. 2: Block Diagram of Automatic Form Process

V. EXPERIMENTAL RESULTS

The fragmented characters are utilized for highlight extraction utilizing PCA and HoG. Execution of these component for various number of classes are contrasted utilizing neural system and back engendering learning. 100 examples for each class are utilized for acknowledgment reason. Neural system parameters are recorded in TABLE 2. Acknowledgment exactness of PCA and HoG for various number of classes.

Table 2: Neural Network Training Parameter

Number of Input nodes	270
No. of Hidden layer	1
No. of Hidden nodes	125
Training epochs	30000
Goal achieved	10e-6
No. of Output nodes	57

Figure 3: Original Skewed Form

Figure 4: Skew Corrected Image

Figure 5: Skew Corrected Image

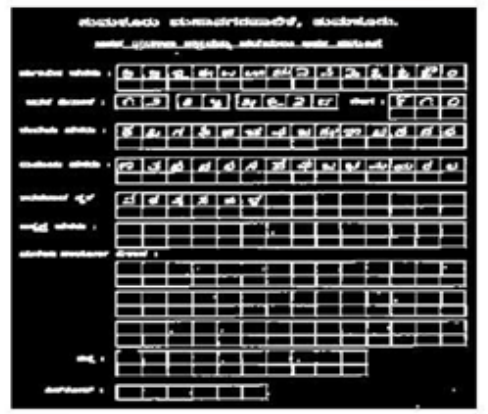

Figure 6: Dilated Form

Figure 7: Hole Filled Image

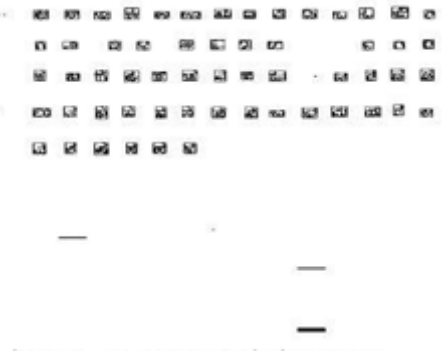

Figure 8: Segmented Characters

VI. CONCLUSION

In this paper, Kannada character acknowledgment with application to programmed structure handling is introduced. Just 57 characters are considered for preparing. Utilizing reasonable pre-preparing systems just transcribed characters are extricated. PCA and HoG are utilized for highlight extraction. Execution of highlights is analyzed utilizing neural system classifier. Consequences of pre-preparing actualized utilizing Visual Studio utilizing OpenCV library are exhibited. Hoard is found to have preferable acknowledgment precision over PCA for substantial number of classes

REFERENCES

1. AfefKacem, AsmaSaidani, AbdelBelaid,"Asystem foranautomaticreadngofstudentinformationsheets",InternationalConferenceonDocumentAnalysisandRecognition,pp1265-1269,2011.
2. M.ThungamaniandP.RamakanthKumar,"ASurveyMethodsandStrategiesinHandwrittenKannadaCharacterSegmentation",InternationalJournalofScienceResearch,Vol1,Issue1,June2012.
3. H.RMamathaandandK.Srikantamurthy,"Morphologicaloperationsan projectionprofilebasedsegmentationofhandwrittenKannadadocument",InternationalJournalofAppliedInformationSystems(IJAIS),Vol4,No.5, October2012.
4. NethravathiB,ArchanaC.P,ShashikiranK,A.GRamakrishnanandVIja Kumar,"CreationofhugeannotateddatabaseforTamilandKannadaOHR",12thIEEEInternationalConferenceonFrontiersinHandwritingRecognition,Nov2010,Pages415-420.
5. RM.K.SinhaandH.N.Mahabala,"MachineRecognitionofDevanagaiScript",IEEETransactionsonSystems,ManandCybernetics,Vol.Smc9,No8, August1979.
6. LianaM.LorigoandVenuGovindaraju,"OfflineArabichandwritingrecognition:Asurvey",IEEETransactionsonPatternAnalysisandmachineIntelligence,Vol28,No.5,May2006.
7. VengatesanK., andS.Selvarajan"ImprovedT-Clusterbased schemefor combinationgenescaleexpressiondata"InternationalConferenceonRadar,CommunicationandComputing(ICRCC),pp.131-136.IEEE(2012).

8. KalaivananM.,andK.Vengatesan.”Recommendationsystembasedonatisticalanalysisofrankingfromuser.InternationalConferenceonInformationCommunicationandEmbeddedSystems(ICICES),pp.479-484, IEEE,(2013).
9. K.Vengatesan,S.Selvarajan:TheperformanceAnalysisofMicroarrayDatausingOccurrenceClustering.InternationalJournalofMathematicalScienceandEngineering,Vol.3(2).pp69-75(2014).
10. KVengatesan,VKaruppuchamy,SPragadeeswaran,ASelvaraj,”FAST ClusteringAlgorithmforMaximizingtheFeatureSelectioninHighDimensionalData”,Volume–4,Issue-2,InternationalJournalofMathematical SciencesandEngineering(IJMSE),December2015.
11. Dayashankar Singh et.al., "Hindi character recognition using RBF neural network and directional group feature extraction technique", International Conference on Cognitive Computing and Information Processing(CCIP), March 2015.
12. N. Sandhya et.al., "Broken kannada character recognition — A neural network based approach", International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), 2016.
13. Nikita Singh, "An Efficient Approach for Handwritten Devanagari Character Recognition based on Artificial Neural Network", 5th International Conference on Signal Processing and Integrated Networks (SPIN), 2018.
14. Sushama Shelke et.al., "Performance optimization and comparative analysis of neural networks for handwritten Devanagari character recognition", International Conference on Signal and Information Processing (IConSIP), 2016.
15. Behnam Alizadehashraf et.al., "Persian handwritten character recognition using convolutional neural network", 10th Iranian Conference on Machine Vision and Image Processing (MVIP), 2017.
16. G. O. Young, “Synthetic structure of industrial plastics (Book style with paper title and editor),” in *Plastics*, 2nd ed. vol. 3, J. Peters, Ed. New York: McGraw-Hill, 1964, pp. 15–64.
17. W.-K. Chen, *Linear Networks and Systems* (Book style). Belmont, CA: Wadsworth, 1993, pp. 123–135.
18. H. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1985, ch. 4.
19. B. Smith, “An approach to graphs of linear forms (Unpublished work style),” unpublished.
20. E. H. Miller, “A note on reflector arrays (Periodical style—Accepted for publication),” *IEEE Trans. Antennas Propagat.*, to be published.
21. J. Wang, “Fundamentals of erbium-doped fiber amplifiers arrays (Periodical style—Submitted for publication),” *IEEE J. Quantum Electron.*, submitted for publication.
22. C. J. Kaufman, Rocky Mountain Research Lab., Boulder, CO, private communication, May 1995.
23. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interfaces(Translation Journals style),” *IEEE Transl. J. Magn.Jpn.*, vol. 2, Aug. 1987, pp. 740–741 [*Dig. 9th Annu. Conf. Magnetics* Japan, 1982, p. 301].
24. M. Young, *The Techincal Writers Handbook.* Mill Valley, CA: University Science, 1989.
25. (Basic Book/Monograph Online Sources) J. K. Author. (year, month, day). *Title* (edition) [Type of medium]. Volume(issue). Available: http://www.(URL)
26. J. Jones. (1991, May 10). Networks (2nd ed.) [Online]. Available: http://www.atm.com
27. (Journal Online Sources style) K. Author. (year, month). Title. *Journal* [Type of medium]. Volume(issue), paging if given. Available: http://www.(URL)

AUTHORS PROFILE

SHAKUNTHALA B S Currently working as Asst.Professor in the Department of ISE at Kalpataru Institute of Technology, pursuing Ph.D degree in Computer Science and Engineering from VTU, Karnataka , India..

Dr. C S Pillai Currently working as Professor in department of CSE at ACS college of engineering, Bangalore, Karnataka , India.