ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2014

Handwritten Character Recognition Mistreatment Multiclass SVM Classification with Hybrid Feature Extraction

A.Rama[1], B.Nagalakshmi[2]

Assistant Professor, Dept. of Master of Computer Application, Bharath University, Chennai, Tamil Nadu, India[1]

Assistant Professor, Department of Computer Science, Bharath University, Chennai, Tamil Nadu, India[2]

ABSTRACT: In this paper, we tend to describe hybrid feature extraction for offline written character recognition. The projected technique could be a hybrid of structural, applied math and correlation options. within the opening, the projected technique identifies the kind and placement of some elementary strokes within the character. The strokes to be hunted for comprise horizontal, vertical, positive slant and negative slant lines–as we tend to observe that the structure of any character are often approximated with the assistance of a mix of straightforward line strokes. The strokes ar known by correlating completely different segments of the character with the chosen elementary shapes. These normalized correlation values at completely different segments of the character offer correlation options. for creating feature extraction additional strong, we tend to add within the second step sure structural/statistical options to the correlation options. The additional structural/statistical options ar supported projections, profiles, invariant moments, endpoints and junction points. This increased, powerful combination of options leads to a 157-variable feature vector for every character, that we discover adequate enough to unambiguously represent and determine every character. Prior, written character recognition downside has not been self-addressed the means our projected hybrid feature extraction technique deals with it. The extracted feature vector is employed throughout the coaching section for building a support vector machine (SVM) classifier. Thetrained SVM classifier is after used throughout the testing section for classifying unknown characters. Experiments were performed on written digit characters and uppercase alphabets taken from completely different writers, with none constraint on style. The obtained results were compared with some connected existing approaches. attributable to the projected technique, the results obtained show higher potency concerning classifier accuracy, memory size and coaching time as compared to those different existing approaches.

KEYWORDS: character recognition, feature extraction, correlation perform, SVM

I. INTRODUCTION

Handwritten character recognition (HCR) is that the pc based mostly identification of written numerals and alphabets. HCR could be a step towards the automation of human interaction with machines. HCR has applications for helping visually-impaired people; for automatic info recording and filtering of written documents; author identification and signature verification etc.. Despite its tremendous scope of application, HCR is a troublesome object classification task as a result of every author has its own means of writing characters and writing fashion varies for one author too.

II. FEATURE EXTRACTION AND CONNECTED WORK

One of the most necessary phases in with success achieving character recognition is the task of feature extraction. Feature extraction stage identifies and extracts varied attributes from characters that facilitate clearly and unambiguously distinguish completely different characters. A range {of completely different|of various} feature extraction ways have been projected in literature in accordance with different character representations. as an example, completely different sets of options have been outlined to best represent character shapes, boundaries, their skeletons and strokes etc. differing kinds of options and ways for character recognition task. Among these ways, there ar applied math feature extractors and structural feature extractors. applied math options take into

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2014

account the arrangement of pel values. Major applied math options used for written character recognition task embody sectionalisation, projections, profiles, and crossings etc. Structural options take into account the pure mathematics and topology of character samples like range of loops, end points, junction points, ratio, sort of strokes and their directions etc. Some feature extraction ways ar {based|based mostly|primarily based mostly} on completely different reworkations such as those based on Fourier transform, rippling rework, central moments, and Zernike moments etc. In [3], the authors describe a sectionalisation based mostly feature extractor to acknowledge written numerals of Indian Kannada script. Authors in [4] acknowledge written numerals mistreatment Fourier descriptors and neural network. In [5], the authors acknowledge Chinese written characters mistreatment gradient and rippling based mostly options. In [6], the authors extract moment based mostly options so as to acknowledge written Arabic letters. They use genetic rule for feature choice and use SVM to assess the classification error for the chosen feature set.

Instead of that specialize in feature vector supported one illustration of a personality, it's a trend currently of combining {different|totally completely different|completely different} sorts of options extracted from different representations of constant character. The advantage of mixing, and harnessing, such completely different forms of options is that it can give wider vary of identification clues to facilitate improve the accuracy of recognition. as an example, Heutte et al. [7] mix completely different applied math and structural options for recognition of written characters. They construct a 124-variable feature vector comprising following seven families of features: 1) intersection of the character with horizontal and vertical straight lines, 2) invariant moments, 3) holes and pouch-shaped arcs, 4) extremas, 5) finish points and junction points 6) profiles, and 7) projections. Aurora et al. [8] mix completely different feature extraction techniques such as intersection based mostly options, shadow options, chain code and curve fitting options for Indian Devnagari language script.

III.PATTERN CLASSIFICATION AND CONNECTED WORK

The second most necessary part in with success achieving written character recognition is that the pattern classification stage. This stage can assign associate unknown character sample to at least one of potential categories by utilizing the data of feature extraction stage. differing kinds of classifiers are often engineered supported the character and sort of knowledge samples and therefore the extracted options. Classifiers used for character recognition downside embody k-nearest neighbor classifier, hidden Andrei Markov model (HMM), support vector machine (SVM), and artificial neural network (ANN) etc. Jain et al. [10] provides a review of applied math pattern recognition techniques. In [11], Pal and Singh train neural network to acknowledge uppercase written characters supported Fourier descriptors of character boundaries as options. In [12], recognition of written alphabets mistreatment neural network and sectionalisation based mostly diagonal options is self-addressed. In [13], Shubhangi and Hiremath acknowledge English written characters and digits by extracting structural small options for SVM classifier. Nasien et al. [14] additionally use SVM classifier to acknowledge written alphabets by using freewoman Chain codes because the options. In [15], Train et al. acknowledge accented written French characters supported a mix of structural and moment options for SVM classifier. In [16], Liu and Nakagawa offer a review of learning ways for nearest neighbor classifiers. [17] and [18] build HMM to acknowledge, severally, offline written Chinese characters and on-line English characters.

IV.PRESENT WORK

In this paper we tend to propose a completely different hybrid feature extraction technique that includes a gaggle of one hundred correlation options aboard with another fifty seven structural/statistical options. Our correlation options ar supported Pearson‟s correlation [19-20] that has been wide applied for the aim of measurement similarity or inequality among the photographs. The worth of correlation constant indicates the extent to that 2 pictures ar similar. Here, request the application of Pearson‟s correlation in an exceedingly completely different means therefore on determine the fundamental elementary strokes in written characters. For this, we tend to work out the correlation constant among completely different character segments and therefore the chosen elementary shapes. we tend to rework the character pictures in frequency domain then we tend to normalize their energy values because it could be a documented reality in signal process theory that the correlation in abstraction domain is merely the multiplication in frequency domain. Shioyama and Hamanaka [extract similar correlation perform based mostly options for the

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2014

matter of Chinese hand-printed character recognition. They but perform their classification supported minimum distance call rule. We, on the contrary, perform final classification supported support vector machine (SVM). the largest challenge, in achieving high accuracy results for SVM classification issues, is the extraction of sturdy options from the info samples. perform is based mostly on power spectral density of character pictures, it's invariant underneath a translational rework and thus will absorb the native variation in hand-printing. during this paper, we tend to take a look at the appliance of this correlation perform based mostly approach to the domain of English written alphabets and numerals. To the simplest of our data, such quick Fourier rework (FFT) based mostly correlation approach has not been nonetheless applied for the classification of English written character samples, although some important work on fuzzy rules based mostly identification of lines and curve strokes within the characters will exist [4].

In our case of at liberty written character recognition downside, these correlation options alone didn't offer satisfactory accuracy for SVM classification. to create the feature vector additional strong, with regard to capability of higher distinctive the characters, we tend to mix correlation perform based mostly options with variety of structural or applied math options. Some structural uncovered options ar finish points and junction points that we tend to add to the correlation options. Finally we tend to add profiles, projections, and moment options to our correlation options as these ar based mostly on binary pictures of characters whereas correlation options ar supported skeletonized characters.

V. PROJECTED METHODOLOGY

Our projected work presents a complete written character recognizer. The system are often split into 3 stages a) pre-process, b) projected feature extraction theme, and c) SVM-based coaching and classification. within the following we are going to describe every of those sub-stages intimately.
These options are extracted from the diluted characters. finish points are those having solely one neighbor, whereas junction points have at least three neighbors. we tend to choose the amount of finish points, the amount of junction points, and therefore the x-y locations of those points because the options to be hold on. Since the amount of finish points and junction points will vary from one character kind to a different, we want some strategy to convert these options into fastened length vector. For this purpose, we tend to use the strategy of [7]. most range of finish points and junction points, call it p, ar noted down from the coaching information, and their average worth with corresponding x-y position is computed. If any character has but p pointes, then empty row in feature vector is stuffed with the typical worth. If throughout testing section, the character happens to possess larger range of points than p, then additional points ar merely discarded. of these options ar normalized in vary [1].

VI. PROJECTION HISTOGRAM

We work out the normalized options as in [7] from accumulative vertical and horizontal projection bar chart of the characters. The coordinate axis of each histograms is divided into eleven equal elements and ten corresponding points on the coordinate axis ar taken because the options.

PROFILES:

We work out normalized options as in [7] from the high, bottom, left and right profiles. when computing the profiles, we tend to notice first-difference profiles by taking the distinction of a degree from its previous one. The maximas in four distinction profiles ar hold on as options. Then the distinction between left and right profiles at 1/3, 2/3, and 5/6 of the character‟s height is recorded as options. Similarly, the distinction between high and bottom profiles at 1/3, 2/3, and 5/6 of the character‟s dimension ar hold on as options.

INVARIANT MOMENTS

The invariant moments live pel distribution around the center of character image. These moments ar invariant to position, size and orientation of the character. we tend to work out seven invariant moments for all the characters and store them as options.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2014

SVM BASED MOSTLY CLASSIFICATION

Once the feature extraction stage is complete, our next section was to make associate intelligent classifier on the extracted feature vector of all the info samples[5][6]. during this analysis we've got chosen the SVM classifier for coaching and classification purpose. SVM could be a 2-class classifier that separates the info samples of 2 categories by computing a maximum-margin boundary between them. The answer for this separating boundary is expressed within the type of a mathematical optimisation downside and it's well- established in SVM literature [29]. In case, the info is nonlinearly severable, SVM makes the info linearly severable mistreatment kernel functions[8]-[10]. A kernel perform maps the input information patterns to some high dimensional area to create the points linearly severable in high dimensional area. Common kernel performs used for classification ar mathematician radial basis function, hyperbolic tangent, polynomial kernel, etc. The separating boundary between the 2 categories is outlined as call boundary and ar known as Support Vectors (SV). These SV verify the separating hyperplane. The binary SVM categoryification downside will be regenerate to multi-class classification by building variety of 2-class SVM classifiers {for completely different|for various} class pairs then taking the ultimate classification call based mostly on different ways such as max-wins strategy, winner-takes-all strategy etc. Max-wins strategy is that the majority-voting call of all the 2- category SVM classifiers. In winner-takes-all strategy, the binary classifier with highest output perform takes the call of classification. Common existing approaches [30-32] for multi-class classification downside ar one-against-one (OAO), one-against-all (OAA), binary tree of SVM and directed acyclic graph (DAG) etc. during this analysis, we tend to have chosen OAO technique for multi-class classification[11].

VII.RESULTS, ANALYSIS AND DISCUSSION

We tested this technique on written characters taken from thirty completely different writers, World Health Organization were allowed to put in writing in their natural vogue. the entire system was enforced in MATLAB. when the pre-processing stage, we tend to extracted a complete of 6092 characters for written uppercase alphabets and 2279 written digits from the scanned documents. information samples were divided into 2 parts: a two-third of knowledge samples was reserved for coaching purpose whereas simple fraction of knowledge samples was reserved for testing purpose. consequently, alphabets coaching information consisted of 4067 characters whereas alphabets testing information comprised 2025 characters. Similarly, digits coaching information consisted of 1857 numerals whereas digits testing information consisted of 922 numerals[12][13]. Feature vectors of dimension 157 were extracted for the coaching information of written characters and numerals. One SVM model was trained on 157 4067 feature matrix of alphabets and another was trained on 157 1857 feature matrix of written digits. SVM parameters on coaching information were fine-tuned mistreatment 3- fold cross-validation. Once the SVM models of written alphabets and digits were trained, we tend to checked performance of the popularity system on reserved testing information sets. Out of the testing information, solely 32/922 digits and 80/2025 alphabets were misclassified. this provides ninety six.5% recognition accuracy on chosen digits information and ninety six recognition accuracy on chosen alphabets information. The system showed 100% accuracy on coaching information of each alphabets and numerals. Its coaching time and memory size of found classifier is way less compared to the opposite 2 approaches[14]. The system has additionally higher recognition rate as compared to different 2 approaches. we tend to more examined the performance of our system on information samples of a new author not originally among the thirty writers on whom the system was trained and tested. show performance of the system on this new author[15]. we tend to ascertained throughout the feature extraction stage that the dilution method generally eliminates necessary character strokes that cause some characters to induce misclassified. The system performance will so be more improved by purification the dilution stage.

VIII.FURTHER ANALYSIS

Our projected hybrid feature extraction technique in conjunction with SVM classifier has shown smart performance on written digits and uppercase alphabets. In future, we tend to will take a look at the performance of projected technique on grapheme alphabets. To acquire satisfactory accuracy on minuscule characters, our hybrid technique would possibly change the window size and form of elementary segments together with, if necessary, some small structural options specific to grapheme characters[16].

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2014

IX. CONCLUSION

A complete offline written character recognition system based mostly on a hybrid feature extraction technique has been conferred. The system comprised 3 main stages, i.e. pre-processing, feature extraction technique, and SVM based mostly training/classification. The projected hybrid feature extraction technique, as experiments unconcealed, established to capture native and international variations in written character designs. The extracted feature vector was a mix of correlation perform based mostly options and some statistical/structural options.

REFERENCES

1.N. Shanthi and K. Duraiswamy, "A novel SVM based mostly written Tamil character recognition system," Springer Journal on Pattern Analysis Applications, Feb 2009.
2.Udayakumar R., Khanaa V., Saravanan T., "Analysis of polarization mode dispersion in fibers and its mitigation using an optical compensation technique", Indian Journal of Science and Technology, ISSN : 0974-6846, 6(S6) (2013) pp. 4767-4771.
3. O. D. Trier, A. K. Jain and T. Taxt, " Feature System‟s performance on digit samples of latest author extraction ways for character recognition: a survey" Pattern Recognition twenty nine (4), 641-662. 1996.
4.Bhuvaneswari B., Hari R., Vasuki R., Suguna, "Antioxidant and antihepatotoxic activities of ethanolic extract of Solanum torvum", Asian Journal of Pharmaceutical and Clinical Research, ISSN : 0974-2441, 5(S3) (2012) pp. 147-150.
5.S. V. Rajashekararadhya and P. V. Ranjan, "Zone based mostly feature extraction rule for written numeral recognition of Kannada script," IEEE International Advance Computing Conference (IACC), Patiala, India, March 2009.
6.Udayakumar R., Khanaa V., Saravanan T., "Chromatic dispersion compensation in optical fiber communication system and its simulation", Indian Journal of Science and Technology, ISSN : 0974-6846, 6(S6) (2013) pp. 4762-4766.
7.Y. Y. Chung and M. T. Wong, "Handwritten character recognition by Fourier descriptors and neural network," IEEE TENCON, Speech and Image Technologies for Computing and Telecommunications, 1997.
8.Sathyanarayana H.P., Premkumar S., Manjula W.S., "Assessment of maximum voluntary bite force in adults with normal occlusion and different types of malocclusions", Journal of Contemporary Dental Practice, ISSN : 1526-3711, 13(4) (2012) pp.534-538.
9.W. Zhang, Y. Y. Tang and Y. Xue, "Handwritten character recognition mistreatment combined gradient and rippling options," International Conference on machine Intelligence and Security, pp. 662-667, Guangzhou, Nov. 2006.
10.Udayakumar, R., Khanaa, V., Saravanan, T., "Synthesis and structural characterization of thin films of sno2 prepared by spray pyrolysis technique", Indian Journal of Science and Technology, ISSN : 0974-6846, 6(S6) (2013) pp.4754-4757.
11.G. Abandah and N. Anssari, "Novel moment options extraction for recognizing written Arabic letters," Journal of engineering science, vol. 5, issue 3, pp. 226-232, 2009.
12. L. Heutte, J. V. Moreau, T. Paquet, Y. Lecourtier, and C. Olivier, " Combining structural and applied math options for the popularity of written characters," Proceedings of thirteenth International Conference on Pattern Recognition, Vienna, Austria, 1996, Vol. 2, pp. 210-214.
13.S. Arora, D. Bhattacharjee, M. Nasipuri, D. K. Basu and M. Kundu, "Combining multiple feature extraction techniques for written Devnagari character recognition," IEEE Region ten Colloquium and third International Conference on Industrial and data Systems, Dec. 2008.
14.Y. Kimura, A. Suzuki, K. Odaka, "Feature choice for character recognition mistreatment genetic rule," IEEE Fourth International Conference on Innovative Computing,Information and management (ICICIC), Kaohsiung , pp. 401-404, Dec. 2009.
15.A. K. Jain, P. W. Duin, and J. Mao, "Statistical pattern recognition: a review," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 1, Jan. 2000.
16.A. Pal and D. Singh, "Handwritten English character recognition mistreatment neural network," International Journal of engineering science and Communication, vol. 1, no. 2, pp. 141-144, July-Dec 2010.
17.T.Nalini ,A.Gayathri,HVS Based Enhanced Medical Image Fusion ,International Journal of Innovative Research in Computer and Communication Engineering,ISSN (Print) : 2320 – 9798 , pp 170-173, Vol. 1, Issue 2, April 2013.
18.S.Thirunavukkarasu, r.K.P.Kaliyamurthie ,EFFICIENT ALLOCATION OF DYNAMICRESOURCES IN A CLOUD,International Journal of Innovative Research in Computer and Communication Engineering, ISSN: 2249-2651, pp 24-29,Volume1 Issue3 Number2–Dec2011
19.K.G.S. VENKATESAN,Planning in FARS by dynamic multipath Reconfiguration system failure recovery in Wireless Mesh Network,International Journal of Innovative Research in Computer and Communication Engineering,ISSN(Online): 2320-9801,pp 5304-5312,Vol. 2, Issue 8, August 2014.
20.G.Michael, An Empirical Approach – Distributed Mobility Management for Target Tracking in MANETs ,International Journal of Innovative Research in Computer and Communication Engineering , ISSN (Print) : 2320 – 9798 , pp 789-794 , Vol. 1, Issue 4, June 2013.
21.G.AYYAPAN , Malicious Packet Loss during Routing Misbehavior - Identification, International Journal of Innovative Research in Computer and Communication Engineering, ISSN(Online): 2320-9801, pp 4610-4613 ,Vol. 2, Issue 6, June 2014