2014 Fourth International Conference of Emerging Applications of Information Technology

Handwritten Bangla Word Recognition using HOG Descriptor

Showmik Bhowmik, Md. Galib Roushan, Ram Sarkar, Mita Nasipuri
Department of Computer Sc. and Engg.,
Jadavpur University,
Kolkata, India.
{showmik.cse, galibr001, raamsarkar, mitanasipuri}@gmail.com

Sanjib Polley[1], Samir Malakar[2]
[1]Department of Computer Sc. and Engg.,
[2]Department of Master of Computer Application,
MCKV Institute of Engineering,
Howrah, India.
{sanjibp6, malakarsamir}@gmail.com

Abstract — The holistic approaches for handwritten word recognition treat the words as single, indivisible entity and attempt to recognize words from their overall shape. In the present work, a novel technique to recognize handwritten Bangla word is proposed. Histograms of Oriented Gradients (HOG) are used as the feature set to represent each word sample at the feature space and a neural network based classifier is applied to classify the word images. On the basis of the HOG feature set, the performance achieved by the technique on a small dataset is quite satisfactory.

Keywords— Holistic word recognition; handwritten words; Histograms of oriented gradients; Bangla script;

I. INTRODUCTION

The growing trend of digitization of books and manuscripts requires an immediate solution for their easy electronic access. This will make the valuable archives to be searchable and usable by the user. A traditional solution to this problem is Optical Character Recognition (OCR) which takes scanned images of paper documents as input and converts them into digital format for computer aided data processing. OCRs have wide range of applications like library and office automation, form and bank cheque processing etc. OCRs can be either online or offline. Online handwriting recognition deals with a data stream which is coming from a transducer while the user is writing. On the other hand Offline handwriting recognition deals with the images of document pages, which have been obtained by scanning the documents. For offline recognition, document pages may be printed or handwritten. Recognition of text is always more complex, in case of handwritten document than the printed or typed one due to the complex nature (like cursive, discrete or mixture of both) of handwritten words. In last few decades, enormous work has been developed in this field and as a result many high quality OCR systems are commercially available in the market for certain languages.

However, for Indian languages, there are no robust OCR engines available, as yet [1]. Different research groups are working on building OCR engines for Indian languages [1], but their performances are not good enough for commercial use. The fundamental challenges in developing an accurate Indic script OCR system are: 1) an extended character set, 2) complex script layout [2], 3) conjoined vowel/consonant modifiers etc. Also to design a good OCR system, one needs to go through several modules like text line and word extraction, character segmentation and recognition etc. For handwritten documents, there are several challenges in every module due to large variations in the writing styles of the individuals. Due to these problems, researchers have begun to focus on holistic word recognition approach. Holistic word recognition is segmentation free approach where a word is considered as a single and inseparable unit and it attempts to recognize words as a whole using their representational features, without any character segmentation. By focusing on recognizing words, instead of characters, better character disambiguation is possible as well as difficulties arise in different modules of the OCR system can be avoided. In the present work, a novel holistic method for handwritten Bangla word recognition system is proposed.

II. PREVIOUS WORK

From literature survey of handwritten word recognition techniques, it is found that, in general, two major approaches are followed by the experts. The first one is *Analytical approach* [3-11, 16, 17], which is a segmentation based approach and the other one is segmentation free or *Holistic approach* [12-15, 18]. In [3], a survey on OCR systems is provided by the authors where a long history of OCR systems is discussed. The classifiers used to recognize characters have evolved over time from Neural Network [5, 6] to Support Vector Machines (SVM) [7]. Errors in classification are typically corrected using post-processing based on character error models [8], dictionaries [9] or statistical language models [5, 10]. OCRs such as the Byblos system [10, 11] uses Hidden Markov models (HMM) to encode the n-gram statistics toward refining character labels. A comprehensive review of Indian scripts recognition is reported in [4]. In the case of Indic-OCRs, despite of significant efforts, their performance is not comparable to that of English.

It is already mentioned that an alternative to OCRs, researchers have started to emphasis on holistic word recognition approaches. In [12], authors provide a survey on the holistic paradigm in human reading and its applications in handwritten word recognition. They show that holistic word recognition is a viable alternative to the popular analytical (segmentation-based) approach to handwriting recognition. In [13], a holistic approach for word recognition is reported using right-left discrete HMM and Kohonen Self-Organizing Feature Map (SOFM) to recognize the Arabic words. In [14], authors proposed a longest-run based holistic feature set that has been used to classify word images belonging to different classes, using a neural network based classifier. To evaluate the technique, few words from the handwritten documents of the *CMATERdb*1.2.1 [15] dataset are used. Frequently occurring English words are manually extracted from the

978-1-4799-4272-5/14 $31.00 © 2014 IEEE
DOI 10.1109/EAIT.2014.43

handwritten pages and the accuracy of the technique is evaluated using a threefold cross-validation method. In [18] a holistic word recognition technique for handwritten historical documents is presented, which is motivated by the fact that for severely degraded documents, segmentation of words into characters often produces poor results. The low quality of the original documents does not allow achieving high accuracy during recognition. The researchers focused on producing transcriptions that will allow successful retrieval of images from historical documents. A document is described using HMM, where words to be recognized represent hidden states. The state transition probabilities are estimated from word bigram frequencies. Word images are represented by fixed length feature vectors, using features ranging from coarse (e.g. word length) to more detailed descriptions (e.g. word profile).

Bangla is one of the most popular languages in the world. Bangla, an Indo-Aryan language of the eastern Indian subcontinent, is evolved from the Bramhi script. This is also the national and official language of Bangladesh and the official language of some of the states of India such as West Bengal and Tripura. Modern Bangla script has 11 vowels and 39 consonant. Apart from vowels and consonants, there are around 280 compound characters in Bangla script [19], which are formed by combining two or more consonants. Despite the several attempts by the researchers, mature research in handwritten word recognition for Bangla script is still a distant dream [16, 17]. This has been our primary motivation of the present work. In this work, Histogram of Oriented gradients (HOG) are used as the feature set, because of its applicability and popularity in the field of object recognition.

III. PRESENT WORK

In the present work, handwritten word samples are collected in A4-size sheets from various persons. After scanning the collected datasheets, word images are cropped. A database for the present work is prepared from such word images. HOG features are then extracted from each word image. Finally, a neural network based classifier is used to recognize them. Fig. 1 shows the entire process of the proposed work.

A. Basic concept of HOG Descriptor

HOG feature descriptors are used in computer vision and image processing for the object recognition purpose. The technique counts the existence of gradient orientation in circumscribed portions of an image. Navneet Dalal and Bill Triggs first reported HOG descriptors in [20]. In this paper, they proposed the scheme for pedestrian detection in steady images. Since then they extended their method for human detection in films and videos, as well as for detection of a variety of common animals and vehicles in static imagery.

The main idea behind the HOG descriptors is that local object appearance and shape within an image can be described by the distribution of intensity gradients or edge directions. The implementation of these descriptors can be achieved by dividing the image into small connected regions, called *cells*, and for each *cell* computing a histogram of gradient directions or edge orientations for the pixels within the *cell*. The combination of these histograms then represents the descriptor (see Fig. 2).

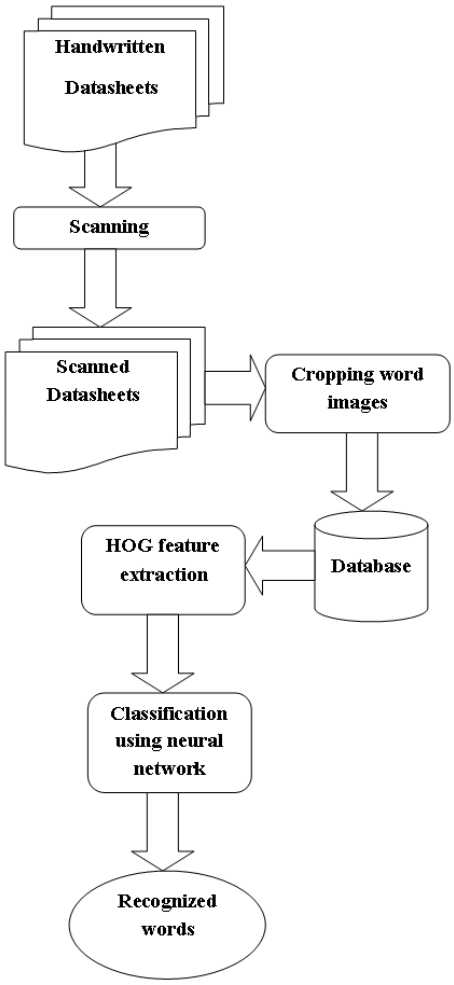

Fig 1: Flow chart of the present word recognition technique

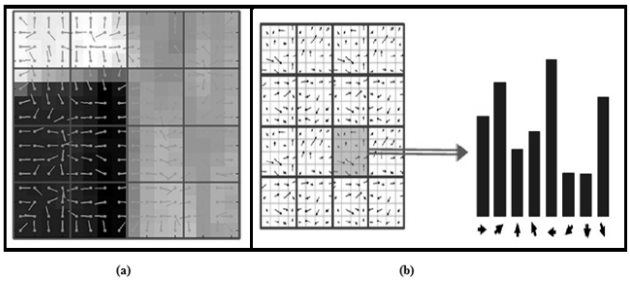

Fig 2: (a) Distribution of intensity gradients or edge directions in an image, (b) histogram computation

For the improvement of the accuracy, the local histograms are contrast normalized. Authors have implemented several normalization schemes for this purpose. The normalization is basically done by grouping *cells* into large spatial blocks. For this purpose, two spatial classes of block geometries are implemented; 1) rectangular or square blocks and 2) circular blocks. Rectangular blocks are partitioned into square or

rectangular *cells*; where as the circular blocks are partitioned into *cells* in log-polar fashion. Each such block is normalized using four different schemes. Let v be an un-normalized vector and $||v||_m$ be its m-norm for m=1, 2 and Є, where Є is a small constant, (a) L2-norm, v → v/ √ ($||v||_2$^2+Є^2). (b) L2-Hys, L2-norm followed by clipping (limiting maximum value of v to 0.2) and renormalized. (c) L1-norm, v→ v/ ($||v||_1$+Є). (d) L1-sqrt, L1-norm followed by square root v→ v/√ ($||v||_1$ +Є), which considers the descriptor vectors as probability distributions.

They have also investigated an alternative centre-surround style *cell* normalization scheme, in which the image is tiled with a grid of *cells* and for each *cell* the total energy in the *cell* and its surrounding region (summed over orientations and pooled using Gaussian weighting) is used to normalize the *cell*. This normalizations result in better invariance to change in illumination or shadowing. The final descriptor is then the vector of all components of the normalized *cell* responses from all of the blocks in the detection window. The authors typically overlap the blocks each so that each scalar *cell* response contributes several components to the final descriptor vector, each normalized with respect to a different block.

B. Feature extraction using HOG Descriptor

The main contribution of the present work is the application of HOG feature for recognition of the handwritten Bangla words. As in the present work, holistic word recognition approach is followed, where a word is considered as an inseparable unit, hence the approach can be thought of as an object recognition procedure. Here, it is assumed that each word sample belongs to a particular class/object. Therefore, it is worth to mention that estimation of HOG feature yields discriminatory result for each object. In this work, said feature extraction process is simplified by considering the non-overlapping windows, which divides the image into small regions, rather than the overlapping blocks and detection window. Fig. 3 shows the feature extraction procedure.

Initially handwritten word images are smoothed with Gaussian filter for the removal of noise. After smoothing, horizontal and vertical gradients of the input image are computed.

Let, the smoothed image is denoted as S(x, y) and the horizontal and vertical components of image gradient at pixel position (x, y) be denoted as $G_x(x, y)$ and $G_y(x, y)$ respectively, which are defined as follows;

$$G_x(x,y)=S(x+1, y) – S(x,y) \quad (1)$$

$$G_y(x,y)=S(x, y+1) – S(x,y) \quad (2)$$

The magnitude M(x, y) and direction D(x, y) of the gradient at pixel (x, y) in the smoothed image are computed as follows;

$$M(x,y)= \sqrt{(G_x^2+G_y^2)} \quad (3)$$

$$D(x,y)=(\tan)^{-1} (G_y/G_x) \quad (4)$$

After computation of magnitude and direction of the gradients, each pixel is pigeonholed in certain category according to its gradient direction. These categories are called *Orientation bin (T)*, which are evenly spaced over 0^0 to180^0 (unsigned gradient) or 0^0 to 360^0 (signed gradient). In the present work, the value of T is chosen as 8. Assigning gradients to the closest bin may result in aliasing noise. To avoid this problem, the gradient magnitude of a pixel can be shared between the two closest bins, which are determined by a linear interpolation in the angular domain.

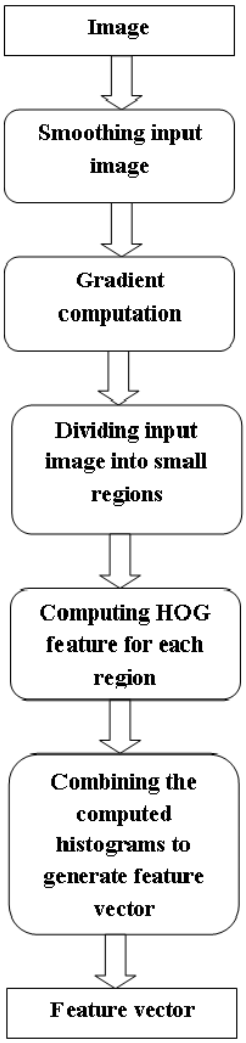

Fig 3: HOG feature extraction process

Let β and (2π/T) –β denote the angles to the two closest bins for a particular pixel (see Fig. 4). Then the contributions of this pixel to the two bins are computed by the following equations:

$$M(x, y) \times [1- T\beta/2\pi] \quad (5)$$

$$M(x, y) \times T\beta/2\pi \quad (6)$$

A non-overlapping window is moved over the input image from left to right direction, which simply divides the image into small sub regions. The magnitudes for similar orientations are accumulated into a histogram for each small sub region (or

for each region under the window). At the end of the process, all those histograms are combined to generate the final feature vector.

In the present work, the input image is divided into 10 small sub regions using the window to get the local information and for each such sub region a histogram with 8 components is computed. Thus from each input image, a feature vector of 80(10x 8) components is obtained. Fig. 5 shows an example of original word sample with corresponding gradients in term of magnitude and direction.

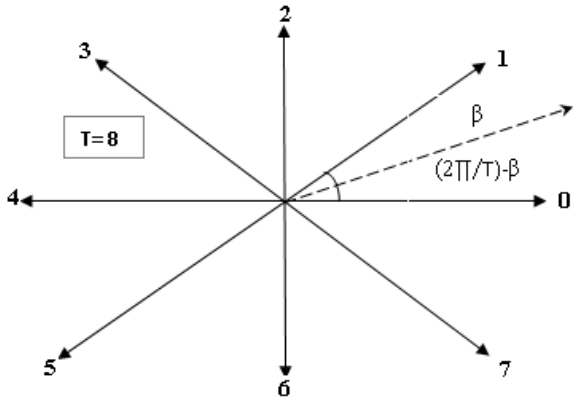

Fig 4: Illustration of angular bins used to extract HOG feature

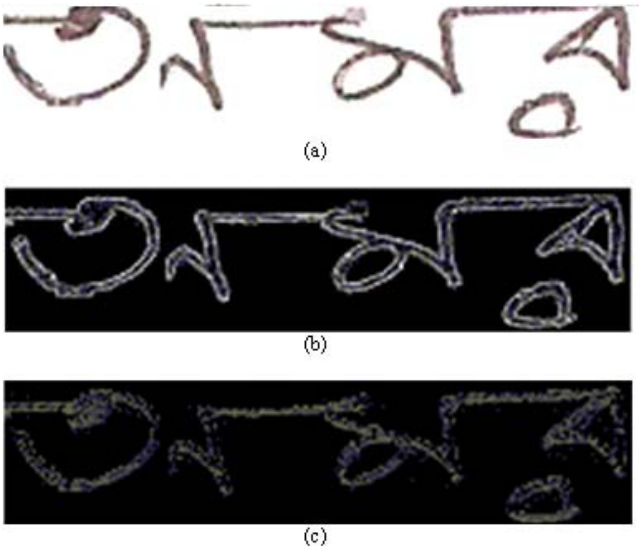

Fig 5: (a) Original word image; Gradients in terms of (b) magnitude (c) direction

IV. CLASSIFIER

In the present experiment, a neural network based classifier, called *Multi-Layer perceptron* (MLP), is used to classify the handwritten words. MLP consists of three different layers; input layer, hidden layer and output layer. Each layer can have certain number of nodes (or neurons) and each node in a layer is connected to all other nodes to the next layer, thus it is also known as feed forward network. The number of nodes in the input layer depends upon the number of attributes present in the dataset whereas the number of nodes in the output layer depends on the number of distinct classes present in the dataset. The appropriate number of hidden layers or the appropriate number of nodes in a hidden layer for a particular problem is hard to decide. But in general, these numbers are selected experimentally. In MLP, the connection between two nodes consists of a weight. During training process, MLP basically learns the appropriate weight adjustment corresponding to each connection. For learning purpose, it uses a supervised learning technique named as *Back propagation algorithm.*

V. EXPERIMENTAL RESULT

In the current work, a machine learning tool, called *weka (Waikato Environment for Knowledge Analysis)* is used. This tool is a collection of several machine learning algorithms, used for data mining, image processing and pattern recognition purpose.

For evaluating the present system, a database containing 1020 handwritten Bangla word images is prepared. This database consists of 20 different classes of words i.e. for a particular word image, 51 different samples are there. Fig. 6 shows some handwritten word samples taken from the said database.

A 3-*fold cross-validation* method is used to evaluate the performance of the system. In each fold, 680 word images are used for training and 340 word images are used for testing. For each fold, an MLP is trained with learning rate (η) = 0.3, momentum term (α) = 0.2 and with different number of neurons in the hidden layer. The detail result is shown in Table 1.

A. Error case analysis

In the present work, though most of the handwritten word images are classified correctly, still there are few cases where word samples are misclassified. Such misclassifications are found in case of *close words.* Two words said to be *close* if they follow almost same pattern of characters (see Fig. 7). As in the present work only gradients are considered as the feature values, therefore, for these *close* words, no significant difference is found in the feature space. As a result, they are classified as the words belonging to the same class.

B. Conclusion

Handwritten word recognition is one of the most challenging concerns in the document images processing. The matter has been addressed by the researchers since last few decades. An enormous work has been done in HWR for different languages like English, French, and Arabic etc. But from the literature survey it is found that only few works have been developed on Bangla script. In the present work, a holistic approach towards the recognition of handwritten Bangla word is proposed.

First, HOG feature set is computed from several local regions of handwritten word images and then a neural network based classifier is used to classify them. Though the current database consists of limited number of word samples, from the experimental result, it can be said that the proposed technique

works satisfactory considering the complex nature and the richness of the Bangla script. In future, we plan to include more number of different word samples in the database. Some misclassifications are found for the words having almost similar character patterns. To cope up these cases, we are also aiming to incorporate some features mainly focusing on the local regions of the words.

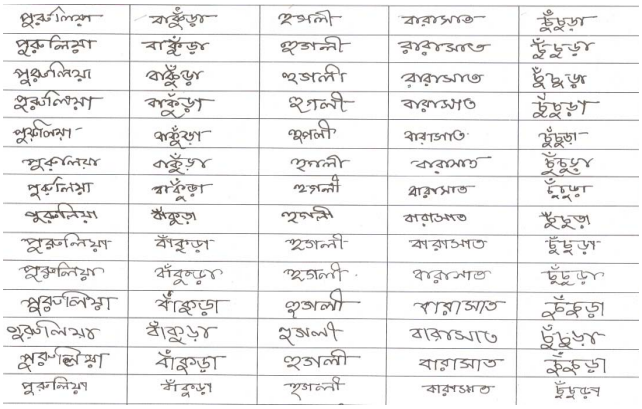

Fig 6: Samples of handwritten Bangla word images

TABLE I. RECOGNITION ACCURACY OF THE PRESENT HANDWRITTEN BANGLA WORD RECOGNITION TECHNIQUE

Fold#	*Number of Training Samples*	*Number of Test Samples*	*Accuracy*
1	680	340	80.29%
2			76.76%
3			**87.35%**

(a) (b)

Fig 7: Two *close* words misclassified by the present word recognition technique

ACKNOWLEDGMENT

The authors are thankful to the Center for Microprocessor Application for Training Education and Research (CMATER) and Project on Storage Retrieval and Understanding of Video for Multimedia (SRUVM) of Computer Science and Engineering Department, Jadavpur University, for providing infrastructure facilities during progress of the work. The current work reported here, has been partially funded by University with Potential for Excellence (UPE), Phase-II, UGC, Government of India.

REFERENCES

[1] V. Govindaraju, S. Setlur. (eds.),"Guide to OCR for Indic Scripts", Springer, Berlin (2009).

[2] S.S.K. Kumar., S. Kumar., V.C. Jawahar,"On segmentation of documents in complex scripts", In: Proceedings of ICDAR, pp.1243–1247 (2007).

[3] G. Nagy, "At the frontiers of OCR", IEEE 80, pp. 1093–1100 (1992).

[4] U. Pal, B. Chaudhuri,"Indian script character recognition: a survey", Pattern Recognit. 37(9), pp. 1887–1899 (2004)

[5] X. Tong, A. D. Evans, "A statistical approach to automatic OCR error correction in context", In: Proceedings of WVLC, pp. 88–10 (1996).

[6] E. Francesconi, M. Gori, S. Marinai, G. Soda, "A serial combination of connectionist-based classifiers for OCR.", International Journal on document analysis and recognition, 3(3) ,pp. 160-168 (2001).

[7] H. Byun, W. S. Lee, "Applications of support vector machines for pattern recognition: a survey". In: Proceedings of 1st International Workshop on Pattern Recognition with Support Vector Machines, pp. 213–236 (2002).

[8] S. Kahan, T. Pavlidis, S.H. Baird, "On the recognition of printed character of any font and size", IEEE Trans. on Pattern Analysis and Machine Intelligence, (2), pp. 274–288 (1987).

[9] G. S. Lehal, C. Singh, R. Lehal, "A shape based post processor for gurumukhi OCR.", In: Proceedings of ICDAR, pp. 1105–1109 (2001)

[10] P. Natarajan, E. MacRostie, M. Decerbo, "The BBN byblos hindi OCR system", Guide to OCR for Indic Scripts, pp. 173–180 (2009).

[11] Z. Lu., R. Schwartz, P. Natarajan,I. Bazzi., J. Makhoul, "Advances in the bbn byblos ocr system", In: Proceedings of ICDAR, pp. 337–340 (1999)

[12] S. Madhvanath and V. Govindaraju, "The role of holistic paradigms in handwritten word recognition". IEEE Trans. on Pattern Analysis and Machine Intelligence, 23(2), pp.149-164 (2001).

[13] M. Dehghan, K. Faez, M. Ahmadi and M. Shridhar, "Handwritten Farsi (Arabic) word recognition: a holistic approach using discrete HMM", Pattern Recognition, 34, pp. 1057-1065, (2001).

[14] A. Acharyya, S. Rakshit, R. Sarkar, S. Basu and M. Nasipuri, "Handwritten Word Recognition Using MLP based Classifier: A Holistic Approach", International Journal of Computer Science Issues, 10(2), pp. 422-427 (2013).

[15] R. Sarkar, N. Das, S. Basu, M. Kundu, M. Nasipuri and D. K. Basu, "CMATERdb1: a database of unconstrained handwritten Bangla and Bangla–English mixed script document image", International Journal on Document Analysis and Recognition, pp. 71–83, (2012).

[16] U. Pal, R. K. Roy, K. Roy and F. Kimura, "Indian Multi-Script Full Pin-code String Recognition for Postal Automation", 10th International Conference on Document Analysis and Recognition, pp. 456-460 (2009).

[17] N. Das, S. Pramanik, S. Basu, P. K. Saha, R. Sarkar, M. Kundu and M. Nasipuri, "Recognition of handwritten Bangla basic characters and digits using convex hull based feature set", In: Proceedings of International Conference on Artificial Intelligence and Pattern Recognition, pp. 380-386 (2009).

[18] V. Lavrenko, T. M. Rath, and R. Manmath. "Holistic word recognition for handwritten historical documents", In: Proceedings of 1st International Workshop on Document Image Analysis for Libraries, pp. 278-287, (2004).

[19] N Das, K Acharya, R Sarkar, S Basu, M Kundu, M. Nasipuri, "A benchmark image database of isolated Bangla handwritten compound characters." International Journal on Document Analysis and Recognition, pp. 1-19, (2014)

N. Dalal, B. Triggs, "Histograms of Oriented Gradients for Human Detection", In: Proceedings of CVPR2005, (1), pp. 886-893(2005).