Hand Written Bangla Numerals Recognition for Automated Postal System

Mostofa Kamal Nasir[1] and Mohammad Shorif Uddin[2]

[1]*Department of Computer Science and Engineering, Mawlana Bhashani Science and Technology University, Bangladesh*
[2]*Department of Computer Sceince and Engineering, Jahangirnagar University, Bangladesh*

Abstract: *Recognition of handwritten Bangla numerals finds numerous applications in postal system automation, passports and document analysis and even for number plate identification. However, the recognition rate requires high and reliable accuracy for practical applications. This paper delineate a robust hybrid system for recognition of handwritten Bangla numerals for the automated postal system, which performed feature extraction using k-means clustering, Baye's theorem and Maximum a Posteriori, then the recognition is performed using Support Vector Machine . Recognition of handwritten numerals, such as postal codes, reveal all kinds of local and global deformations: distortions, different writing styles, thickness variations, wide variety of scales, limited amount of rotation, added noise, occlusion and missing parts. This paper shows that the proposed method is better than other system.*

Keywords - *K-means clustering, Bayes' theorem, MAP, PCA, SVM and OCR.*

I.   Introduction

Optical character recognition (OCR) is an automatic reading of optically sensed document text materials to translate human-readable characters to machine-readable codes. OCR system for printed Bangla numerals, the fourth most popular script in the world [1]. About 200 million people of Eastern India and Bangladesh use Bangla as language. Postal automation is a topic of research interest for last two decades. System development towards postal automation for a country like Bangladesh is more difficult than that of other countries because of its multi-script behavior. Some people write the destination address part of a postal document in two language scripts (i.e. Mixing Bangla and English). Unfortunately, researches on Bangla numeral recognition are not sufficient so far, in particular on handwritten issue. Some Papers on printed Bangla numeral recognition have been reported in past years [1-3]. But there are few researches on handwritten Bangla numeral recognition. Professor Pal. have done some exploring work for the issue of recognizing handwritten Bangla numerals [2-4]. Another well known research work is presented in [5]. In this paper, we propose a knowledge based system for handwritten Bangla numerals detection and shows its implementation of postal automation. This paper also shows the comparison between existing methods and the proposed method.

The rest of the paper is organized as follows: Existing Method is discussed in Section 2. In Section 3, the flowchart of the hybrid system is showed. The proposed method is described in Section 4. Section 5 presents Result and Comparison with Existing System which is followed by conclusions and future works in Section 6.

II.   Existing Method

The handwritten digit recognition application is a machine vision task. The input consists of black or white pixels. The digits are usually well separated from the background and there are only ten output categories. Yet the problem deals with objects in a real two dimensional space and the mapping from image space to category space has both considerable regularity and considerable complexity. The problem has added attraction because it is of great practical value. The database used to train and test the network contains handwriting of many different people. The choice of an appropriate data representation is a crucial point when solving a classification task, either with a trainable or with a nontrainable classifier. Therefore, the original input representation is usually transformed into a higher-level data representation by using human expertise for designing appropriate preprocessing operations. However, besides mere recognition rates, there are other factors which may influence the choice of the data representation, such as the need for high computation speed or hardware limitations, favoring data representations which do not require complicated and time consuming preprocessing. In many cases, this precludes the use of structured data representations. Therefore, the first data representation we use for the subsequent classification is a simple pixel representation. For a given data representation, an optimal hyper surface separating the classes in the N dimensional input space in the best possible way might be found if the underlying probability distributions were either known or estimated accurately. In the case of handwritten digits, it seems therefore natural to use some a priori knowledge about the recognition task in order to transform the low level information of the pixel images into a data representation of

higher level. Two possibilities arise: either the a priori knowledge is used to design preprocessing operations which are carried out explicitly before classification, or it is used to constrain the general architecture of the classifier, which results in a classifier specialized in the recognition task at hand. I have chosen the first possibility. Digits, whether handwritten or typed, are essentially line drawings, i.e. one-dimensional structures in a two-dimensional space. Therefore, local detection of line segments seems to be an adequate preprocessing. For each location in the image, information about the presence of a line segment of a given direction is stored in a feature map. Thus the second data representation used for the subsequent classification consists of one feature map for each detected direction of line segments. The hope is that, with this data representation, the examples of a given class are less dispersed in input space than with the simple pixel representation.

There are several existing methods for handwritten Bangla numeral recognition such as IRPCA, PCA+SVM, KPCA+SVM, KPS, Integrated system (integrated with IRPCA, PCA+SVM and KPS). Among the existing system IRPCA, PCA+SVM, KPCA+SVM has less response time than the proposed system but their recognition rate is poor. KPS has a good recognition rate but not better than the proposed method and has larger response time. Integrated system (integrated with IRPCA, PCA+SVM and KPS) has the best recognition rate compared to others (except the proposed method) but it has large response time which is not desirable.

III. The Hybrid System

In the hybrid system, at first, preprocessing is performed, feature extraction using k-means clustering and Bayes' theorem and Maximum a Posteriori(MAP) is performed, then recognition is performed using support Vector Machine(SVM). This research shows that the proposed method is better than other system. The block diagram of the proposed system is given below in a fig-1.
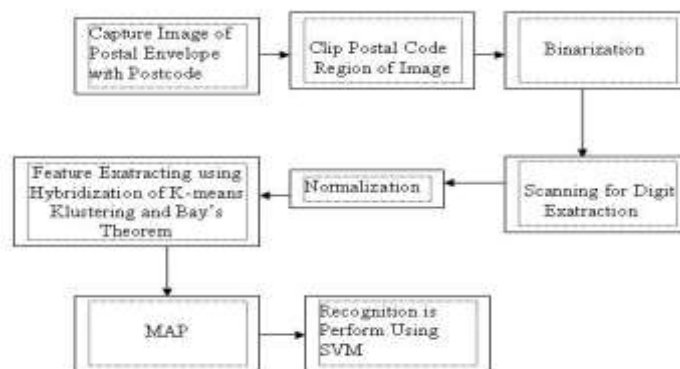

Fig-1: Flowchart for the Proposed System

IV. Proposed Method

4.1Preprossing

This section briefly discusses the preprocessing of data representation. All the preprocessing steps shown in the figure. The goal of the preprocessing is to isolate digits from a captured image. The most difficult part of preprocessing is the automatic segmentation of the zip codes. To avoid this problem we used standard postal envelope of the Bangladesh postal department where there is a four digit box at the lower left corner to write the postal code. The total image was first converted to pgm format and then applying a threshold further converted to a two level gray image. Then connected parts were separated to extract digits. In this method some problem occurred, like, some spots or noise on envelop were found which was separated. To avoid that problem flood fill algorithm was used. The process is discussed below.

4.1.1 Capture Image

The handwritten bangle postal code is first captured using web camera and stored in JPG format. For this project the standard postal envelopes of Bangladesh postal depth are used. In this type of envelop four digits postal codes are written in the box in lower right corner. The digits were written by 75 different people. Some scanned digits were also taken which were written by different people also. In a total 75X4= 300 different digits are collected. These digits vary in shape and size. When capturing images with a web camera a fixed distance was maintained which was 7 cm from the test envelop. The resolution of the camera was 240X320 and the image size is 6.04 KB. The images were in JPEG format when captured with a camera and it is converted to pgm (portable gray map) format for simplicity of manipulation. After converting to pgm format, the image has 256 gray levels from 0 to 255 where 0 indicates a black point where 255 indicate a white point.

4.1.2 Clipping Image

The lower right part of the image is the region of interest where the postal digit is written. The captured images are clipped from 130 pixel values in the x axis and 160 pixel values from the y axis. Thus only the boxes containing handwritten digits are taken. This part has a resolution of 190X80. This image is binarized using a threshold value 102. All pixels having a value greater or equal to 102 was leveled at 1 and all values below 102 was leveled as 0 [6]. Thus the file contains only two levels. After binarization some undesired points were found, which was eliminated using flood fill algorithm. Flood fill, also called seed fill, is an algorithm that determines the area connected to a given node in a multi-dimensional array. The algorithm can be sped up by filling lines. Instead of pushing each potential future pixel coordinate into the stack, it inspects the neighbor lines (previous and next) to find adjacent segments that may be filled in a future pass; the coordinates (either the start or the end) of the line segment are pushed on the stack. In most of the cases this scan line algorithm is at least an order of magnitude faster than the per-pixel one. This algorithm finds region where less than 9 pixels are connected together. It is analyzed that in the examined region any digit contains more than nine digits connected together.

4.1.3 Segmentation

The individual digit is segmented by horizontal and vertical scanning. Horizontal scanning found the region of interest to extract. Vertical scanning segmented individual digits. After extraction 4 individual digit images are found. These images are different in size. Again the images vary in pixel width. So, they are made 1 pixel width thin using thinning algorithm. Then thin images are normalized in 11X14 pixels. These pixels are ready for Feature extraction

4.1.4   Horizontal Scan

Clipped image was scanned horizontally first [7]. In this scanning the first black pixel indicated first row having any digit. Here the region of interest begins. From this location horizontal scanning is done until any such row is found where all pixels are white. That row indicates the ending of the region which contains digits. Thus with horizontal scanning the handwritten region is separated from unwanted area.

4.1.5 Vertical Scan

At the beginning of a region found from horizontal scanning vertical scanning is started. It scans through the y-axis of the region [7]. When scanning discovers any black pixel, it assumes it to be the beginning of any digit. And when gets any all-white column in the way of scanning it thinks it to be the end of that digit. Then search for another black pixel scanning vertically to find the next digit. Thus in four cycle four digits are extracted.

4.1.6 Thinning

Thinning is a morphological operation that is used to remove selected foreground pixels from binary images, somewhat like erosion or opening. It can be used for several applications, but is particularly useful for skeletonization. In this mode it is commonly used to tidy up the output of edge detectors by reducing all lines to single pixel thickness [8]. Thinning is normally only applied to binary images, and produces another binary image as output. The thinning operation is related to the hit-and-miss transform, and so it is helpful to have an understanding of that operator before reading on.
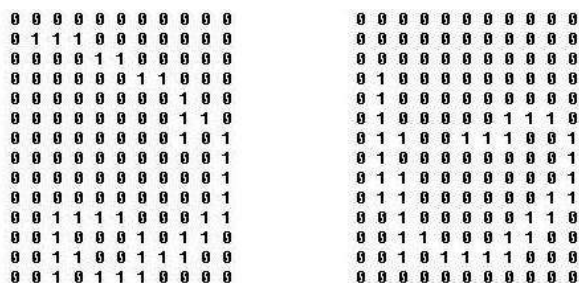

Fig-2: Thinning of the image

4.1.7 Dilation

Dilation is one of the two basic operators in the area of mathematical morphology, the other being erosion. It is typically applied to binary images, but there are versions that work on grayscale images. The basic effect of the operator on a binary image is to gradually enlarge the boundaries of regions of foreground pixels (i.e. white pixels, typically). Thus areas of foreground pixels grow in size while holes within those regions

become smaller[9]. The dilation operator takes two pieces of data as inputs. The first is the image which is to be dilated. The second is a (usually small) set of coordinate points known as a structuring element (also known as a kernel). It is this structuring element that determines the precise effect of the dilation on the input image.

```
1 0 0 0 1 1 1 1 0 0 0          1 1 1 1 0 0 0 0 0 0 0
0 0 1 1 1 1 1 1 1 0 0          0 1 1 1 1 0 0 0 0 0 0
0 1 1 1 0 0 1 0 1 1 0          0 0 0 1 1 1 0 0 0 0 0
1 1 1 0 0 0 0 0 1 1 1          0 0 0 0 1 1 0 0 0 0
1 1 1 0 0 0 0 0 1 1 1          0 0 0 0 1 1 0 0 0 0
1 1 1 1 1 0 1 1 1 1 1          0 0 0 0 1 1 0 0 0 0
0 0 1 1 1 1 1 1 1 1 0          0 0 0 1 1 0 0 0 0 0
0 0 0 1 1 1 1 1 0 0          0 0 1 1 1 0 0 0 0 0 0
0 0 1 1 1 1 0 1 1 1 0          0 0 1 1 1 0 0 0 0 0 0
0 0 1 1 1 0 0 0 1 1 1          0 1 1 1 1 0 0 0 0 0 0
0 0 1 1 0 0 0 0 0 1 1          0 1 1 1 1 0 0 0 0 0 0
0 0 1 1 0 0 0 1 1 1          0 0 1 1 1 1 1 1 0 0
0 0 0 1 1 0 1 1 1 1 1          0 0 0 1 1 1 1 1 1 1
0 0 0 0 1 1 1 1 1 1 0          0 0 0 0 0 0 0 1 1 1
```
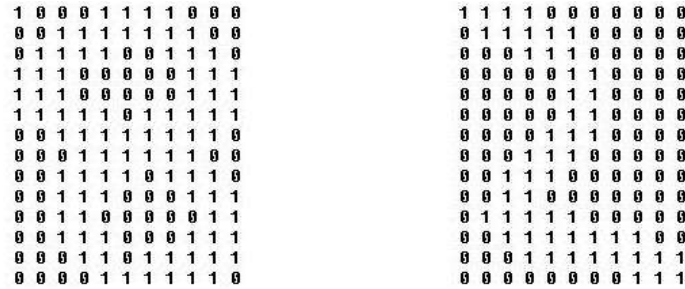
Fig-3: Dilation of Bangla Digit

To compute the dilation of a binary input image by this structuring element, we consider each of the background pixels in the input image in turn. For each background pixel (which we will call the input pixel) we superimpose the structuring element on top of the input image so that the origin of the structuring element coincides with the input pixel position. If at least one pixel in the structuring element coincides with a foreground pixel in the image underneath, then the input pixel is set to the foreground value. If all the corresponding pixels in the image are background, however, the input pixel is left at the background value. Dilation is the dual of erosion i.e. dilating foreground pixels is equivalent to eroding the background pixels

4(a)   4 (b)

Fig-4: (a) original image obtained from post card (b) preprocessed image

4.2 Region Probability Estimation and Region Classification

Step 1: Perform anisotropic smoothing using constrained PDE. The equation is given below

$$\forall_i = 1, \ldots, n, \qquad \frac{\partial I_i}{\partial t} = trace(TH_i) + \frac{2}{\pi} \nabla I_i^T \int_{\alpha=0}^{\pi} J_{\sqrt{T_{a_\alpha}}} \sqrt{T_{a_\alpha}} \, d\alpha \qquad (1)$$

Where $a_\alpha = (\cos\alpha \ \sin\alpha)^T$, and $J_{\sqrt{Ta\alpha}}$ stands for the Jacobian of the vector field $\Omega \to \sqrt{Ta\alpha}$

Step 2: Cluster the pixel into three regions (numerals and background) using K-means clustering method and calculate the mean value(μc) and standard deviation for each class.

Step 3: Calculate the probability density function using the following equation:

$$\Pr(V(x) = v \mid x \in c) = \frac{1}{\sqrt{2\pi}\sigma_c} \exp\left(-\frac{(v-\mu_c)^2}{2\sigma_c^2}\right) \qquad (2)$$

Where V (x) is the intensity value of the pixel at position x = (x, y). We assume that the prior probability that a particular pixel will belong to a certain class is equal and homogeneous, i.e., $\Pr(x \in c)$ is the same for each class and for all pixels.

Step 4: Apply the Bayes' rule to calculate the posterior probability that a given pixel belongs to a particular class, given its intensity:

$$\Pr(x \in c \mid V(x) = v) = \frac{\Pr(V(x) = v \mid x \in c)\Pr(x \in c)}{\sum_\gamma \Pr(V(x) = v \mid x \in \gamma)\Pr(x \in \gamma)} \qquad (3)$$

After applying Region Probability Estimation and Region Classification we get the images shown in fig 2. below

5(a)   5(b)

Fig-5: (a) preprocessed image, (b) resultant image after Region Probability Estimation and Region Classification

4.3 Recognition of numerals

Recognition of numerals is done by Support Vector Machine (SVM). The input of the SVM is the output of region probability estimation and region Classification. SVM is described below:

$$f(x) = \text{sgn}\left(\sum_{i=0}^{l} y_i \alpha_i k(x_i x) - b\right) \tag{4}$$

$$\text{sgn}(u) = \begin{cases} 1 & for \quad u > 0 \\ -1 & for \quad u < 0 \end{cases} \tag{5}$$

The appeal of SVM lies in their strong connection to the underlying statistical learning theory. According to the structural risk minimization principle, a function that can classify training data accurately and which belongs to a set of functions with the lowest capacity (particularly in the VC-dimension) will generalize best, regardless of the dimensionality of the input space. In the case of a canonical hyper plane, minimizing the VC-dimension corresponds to maximizing the margin. As a result, for many applications, SVM has been shown to provide a better generalization performance than conventional techniques. The training data points can be expressed as *(x1, y1), (x2, y2), . . . , (xl, yl)(xi ϵ Rn; yi ϵ R is the target value)*, SVM approximates the function using the following form:

$$f(x) = sgn(w.\Phi(x) + b) \tag{6}$$

where Φ (x) represents a high-dimensional feature space which is nonlinearly mapped from the input space x. The coefficients w and b are estimated by minimizing the regularized risk function . Support vector machine for a pattern recognition problem can be formulated as the quadratic optimization problem [10].

V. Result and Comparison with Existing System

Total 180 digits, including some scanned digits were taken as sample digits. All of was extracted correctly. This the Rate of success is 100% in the preprocessing phase.

Table 1: Result of the Experiment.

No. of sample digits	300
No. of Extracted Digits	298
No. of error Extraction	2
Rate of Success	99.33%
Rate of error	0.67%

The performance of an OCR system depends on two facts namely response time and recognition rate. Table-2 shows a recognition rate comparison and Table-3 shows the response time comparison.

Table 2: Recognition rate comparison

Approach	0	1	2	3	4	5	6	7	8	9	Avg
IRPCA	93.55	89.17	90.53	86.27	87.70	90.08	94.20	98.79	98.02	94.44	92.27
PCA+SVM	97.01	91.12	70.38	82.59	93.33	93.14	77.23	91.27	90.67	83.37	87.01
KPCA+SVM	97.08	92.11	71.25	81.85	93.33	92.84	76.37	91.36	90.66	84.10	87.10
KPS	99.06	92.52	93.58	91.95	97.88	97.25	93.14	99.51	98.42	90.64	95.36
Integrated system	98.87	94.41	92.56	91.25	95.56	95.25	93.02	98.36	97.89	93.25	95.06
Proposed method	98.09	95.75	94.56	96.36	99.32	96.88	95.96	97.25	96.99	96.83	96.80

Table 3: Response time comparison

IRPCA(s)	KPS(s)	Integrated system	Proposed System
0.105	0.23	0.36	0.21

From the above two tables it is clear that the response time is considerable and the recognition rate of the proposed method is better than any other system.

VI. Conclusion

This paper proposes an efficient OCR system for handwritten Bangla numeral recognition. The proposed method is based on preprocessing, k-means clustering, Bayesian theorem and SVM. An efficient recognition system for handwritten Bangla numerals has been developed. The experimental results confirm the

effectiveness of the proposed approaches. We employ the newest hybridization strategy in the system for improving recognition performance. The recognition result achieved by the hybrid system is more reliable than that by one method alone or the integrated method. Experiments also demonstrate that the response time of the hybrid system is acceptable. It is obvious that hybridization of different approaches will improve the performance of the whole recognition system. In our future work, we will include more recognition approaches in our hybrid system for achieving better recognition ability in the case that the response time is acceptable. We will also work not only for better recognition for postal code but also for the whole address part automated

References

[1] B. B. Chaudhuri, U. Pal, "A Complete Printed *Bangla* OCR System", *Pattern Recognition*, Vol. 31, No. 5, pp. 531-549, 1998
[2] U. Pal, B.B. Chaudhuri, Indian script character recognition: a survey, Pattern Recognition 37, pp:1887–1899, year:2004
[3] U. Pal, On the development of an optical character recognition (OCR) system for printed Bangla script, Ph.D. Thesis, 1997.
[4] U. Pal, B.B. Chaudhuri, Automatic recognition of unconstrained offline Bangla hand-written numerals, in: T. Tan, Y. Shi, W. Gao
[5] (Eds.), Advances in Multimodal Interfaces, Lecture Notes in Computer Science, vol. 1948, Springer, Berlin, pp. 371–378,year: 2000.
[6] K. Roy, S. Vajda, U. Pal, B.B. Chaudhuri, A system towards Indian postal automation, in: Proceedings of the Ninth International
[7] Workshop on Frontiers in Handwritten Recognition, (IWFHR-9 2004), pp. 580–585, year: October 2004.
[8] U. Pal, A. Belad, Ch. Choisy, Touching numeral segmentation using water reservoir concept, Pattern Recognition Lett. 24, 261–272, year: 2003.
[9] YingWen, Yue Lu, Pengfei Shi, "Handwritten Bangla numeral recognition system and its application to postal automation ". Pattern Recognition 40, pp:99 – 107. year:2007.
[10] J.-X. Dong, A. Krzyzak, C.Y. Suen, Fast SVM training algorithm with decomposition on very large data sets, IEEE Trans. Pattern Anal.