A Neural Network based Method for Recognition of Handwritten English Pitman’s Shorthand

Janmaya Kumar Mishra
Department of Computer Science & Engineering, Himalayan University, Arunachal Pradesh, India

Khursheed Alam, Ph.D
School of Engineering & Technology, Sharda University, Greater Noida, U.P, India

ABSTRACT
Pitman Shorthand method of documenting is normally practiced by stenographers to take dictation at speaking speed invented by Sir Isaac Pitman in 1837. Special graphical symbols are used in this method of representing phonetic compositions of the dictated text for certain interval. This shorthand representation itself is a compressed and encrypted format of the English text. The Pitman Shorthand Language (PSL) has the practical advantage of high speed of recording, more than 120–200 words per minute, because of which it is universally acknowledged. This recording medium has its continued existence in spite of considerable developments in speech processing systems, which are not universally established yet. In this method of documenting, speech is directly converted into phonetic strokes, where each phonetic stroke is a composition of consonants and vowels. These consonants are called consonant primitives or simply primitives. In order to exploit the vast transcribing potential of PSL a new area of research on automation of PSL processing is conceived. In this work, the Neural Network Classifier for recognition of PSL at word level is presented. All the activities such as preprocessing of data (image cropping, resizing and implementations) have been carried out using MATLAB. As per this experiment, 94% recognition accuracy is achieved.

General Terms
Pattern Recognition, Neural Network, Feature Extraction, Classification.

Keywords
Pitman's Shorthand, Neural Network Classifier, Hierarchical Centroid

1. INTRODUCTION
Automatic analysis of hand written symbols or characters is one of the human requirements that have been a subject of intensive research for the last few decades and it is still far from the perfection. Advantages of the practical ability of PSL are universally acknowledged and this recording medium continues to exist in spite of considerable developments in Speech Processing Systems (SPS), which are not universally established yet. Pitman is phonetic: it records the sounds of speech rather than the spelling. Vowel sounds are optional and are written with small dots, dashes or other shapes next to the main strokes. This helps increase writing speed because most words can be identified from their consonants only. The thickness, length and position of the strokes are all significant. There are many special abbreviations and other tricks to increase writing speed.

The present work is concerned to substitution of equivalent English words, for the phonetic compositions of the transcribed words in the process of automatic generation of English text from existing PSL document. The present work is the development of a method for recognition of handwritten PSL words [1], thus facilitating automation of English text generation from written PSL written documents. In this research area, there are two approaches available. Online processing and another is Offline processing. In online processing using tablet and pen one can take dictation which gets automatically converted to equivalent English text. Whereas in offline processing the PSL documents are scanned and subjected to text generation. The Figure-1 shows few strokes which includes Down strokes, Horizontal Strokes and Up Strokes. Straight-line primitives and curved primitives (arcs) make the fundamental set of primitives. These primitives are also characterized and identified with their thickness, i.e. strokes may be thin or thick. Hook-structured primitives are used as consonants.

Fig 1.: Pitman’s strokes

The basic primitive patterns, as can be seen in figure 1, are straight lines, arcs and hooks. Thus one has to choose a technique that recognizes these three types of fundamental shapes accurately. A method was proposed to identify fundamental shapes in a stroke using structural feature extraction from PSL words.

2. RELATED WORKS
There are so many researchers worked in various pattern recognition systems related to different hand written languages including regional languages as well as hand written shorthand pattern recognition system. Not enough research has been found in recognition of PSL document. So, summary of following works are explained in this section..
P. Nagabhushan and et al. [2] described an automatic recognition of the strokes of the PSL at word level. A pen-down to pen-up sequence makes a stroke, which is a composition of primitives. The words are separated based on pen-down and pen-up points. The features that form a word (a stroke) are grouped first. Next, primitives and their sequence are identified and passed to a recognizer which identifies the word. A tangent-based vector through the contour of a stroke identifies the consonant primitives. Any other marks close to the stroke but not associated with the contour of a stroke represent the vowel markers. The proposed system implemented and tested the algorithm using 'C' language, used 256-colour images for thick and thin primitives to be differentiated based on their gray levels. A corner point may not be identified even after traversing a standard number of pixels along the primitive from the previous corner point. Detection of consonant primitives and vowel markers in a word by using tangent-based feature vector approach is able to recognize most handwritten PSL words.

Yang Ma et al. [3] presented a recognition system for PSL which includes the followings four major procedures: segmenting and classifying consonant outlines; classifying vowels and diphthong symbols according to their shape characteristics; classifying shortforms with Hausdorff distance; and dealing with the problem of smooth junction with a new machine compatible shorthand. An experiment on 461 outlines each containing one sooth junction shows that the recognition rate achieved was 93.41%.

In a work by Leedham, C. G. and Downton, A. C. [4] Pitman shorthand outlines can be split into two classes of characters, shortforms (comprising over 90 of the most frequently used words and phrases in the English language) and vocalised outlines which can represent any word pseudo-phonetically. The shortforms represent as much as 50% of normal shorthand and are recognised directly using a dynamic programming with template matching technique with typical recognition accuracy of over 90%. Vocalised outlines are recognised using a syntactic method which interacts with a knowledge source derived from analysis of a large number of shorthand outlines.

S. M Htwe, et al.[5] proposed segmentation and recognition of phonetic features in Pitman shorthand. Approaches to the recognition of consonant outlines, vowel and diphthong symbols and shortforms, which are different components of Pitman shorthand, are presented. A new rule is introduced to solve the issue of smooth junctions in the consonant outlines which was normally the bottleneck for recognition. Experiments with a set of 1127 consonant outlines, 2039 vowels and diphthongs and 841 shortforms from three shorthand writers have demonstrated that the proposed solution is quite promising. The recognition accuracies for consonant outlines, vowels and diphthongs, and shortforms achieved 75.33%, 96.86% and 91.86%, respectively. From the evaluation of 461 outlines with smooth junction, the introduction of the new rule has a great positive effect on the performance of the solution. The recognition accuracy of smooth junction improves from 37.53% to 93.41% given a writing time increase of 14.42%.

Leedham, C. G. and Andy C. Downton [6] described a number of evaluation experiments designed to establish the potential of Pitman’s handwritten shorthand as an input for computer transcription to text. Such a system would have applications in verbatim reporting, the electronic office, and as an aid for the deaf. The experiments described compare the performance of a proposed computer transcription system for shorthand with the benchmark performance obtained using human transcription. In addition, measurements on typical Pitman shorthand data are used to estimate potential performance limits. It is concluded that the poor overall performance of the proposed computer transcription system is due to a combination of three factors: first, the simplified nature of the recognition algorithms used compared with the knowledge-based techniques used by human shorthand readers; secondly, ergonomic deficiencies of the data input devices used; and finally, writer’s lack of familiarity with the system and its capabilities. A proposed strategy for improving the performance of the system by attention to the first two of these deficiencies was given.

3. DATA PREPROCESSING AND IMPLEMENTATION DESIGN

A data set created for the research is depicted in figure 2. This research deals with isolated handwritten PSL words. A total of 675 images are scanned, 75 images per PSL word.

These selected RGB images were cropped in order to remove unwanted information from each image and were resized into 300x300 RGB images. Before feature extraction and classification, these RGB images were converted to grey scale image for speed up the processing.

Feature extraction process used for dimensionality reduction of scanned images. As input images are too large for processing, so to process these images in desirable time period, features are extracted from the images. Feature extraction is chosen in such a way that image information must be retained. Feature extraction is an essential pre-processing step to pattern recognition and machine learning problems. It is often decomposed into feature construction and feature selection.

Feature extraction [7] involves simplifying the amount of resources required to describe a large set of data accurately. When performing analysis of complex data one of the major problems stems from the number of variables involved. Analysis with a large number of variables generally requires a large amount of memory and computation power or a classification algorithm which over fits the training sample and generalizes poorly to new samples. Feature extraction is a general term for methods of constructing combinations of the variables to get around these problems while still describing the data with sufficient accuracy.

The feature extraction method used in this research is Hierarchical Centroid[8]. It uses the centroid method of finding the centroid of the image, through the centroid image is partitioned into two different zones, left and right zones. Iteratively this method was performed at most seven times then we get 124 features out the image. The end result of the extraction task is set of features, commonly called a feature vector and feature vector constitutes a representation of the image.

Extracted features were used as input for classification engine. Classification techniques are helpful to recognize the English shorthand. There are a number of classification techniques available as described in [9-11]. Classification is identifying inputs to a set of class on basis of training data set. The scope of image classification is to identify and convey meaning. Image classification is the most important part of digital image analysis. Classification techniques generally aim to provide a reasonable answer for all possible inputs and to perform “most likely” matching of the inputs, taking into account their statistical variation. In this work, Neural
Network Pattern recognition tool (NPRTOOL) was used in recognizing the English shorthand words.

NPR Tool leads in solving a pattern recognition classification problem using a two-layer feed-forward pattern net network with sigmoid output neurons. Fifty neurons were used on the hidden layer. The network had eight output neurons because there are eight target values associated with each input vector. Pattern recognition networks are feed forward networks that can be trained to classify inputs according to target classes. The target data for pattern recognition networks should consist of vectors of all zero values except for a ‘1’ in element i, where i is the class they are to represent.

For the purpose of classification, the input data were divided into two sets: training (includes validation) and testing.

Training data was used for adjusting the weight and biases. Validation was used to decide when to stop the training process, to avoid over fitting which is a situation where the network memorizes the training data, rather than training the low that governs them. Testing data was used to measure the performance of the trained network. It is important that this data did not participate in the training process.

Training set contains 60% of the original data, 20% for validation for testing, the rest 20% were used as a completely independent test of network generalization.

All the characters have undergone different morphological operations using MATLAB functions. The pre-processing steps used in this research are as below:

Step 1: Isolated word extraction
Step 2: Conversion into the negative of the word
Step 3: Thinning the image
Step 4: Image cropping and resizing

All the experiments have been carried out using MATLAB 7 on a standard Core i7 computer with Windows 8 as operating system.

4. IMPLEMENTATION

In this research, the recognition process has been carried out once the network has been established. The network has been created by providing the target vectors. Stable equilibrium points at these target vectors are stored in the network. Hence, there is no need to perform iterative training on it. That is because Neural Network learn patterns in a one-shot style.

### Table 1: Results obtained from the experiment

<table>
<thead>
<tr>
<th>PSL Words</th>
<th>The</th>
<th>Quick Brown</th>
<th>Fox</th>
<th>Jumps</th>
<th>Over</th>
<th>Lazy</th>
<th>Dog</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classified Words</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>6</td>
<td>7</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Misclassified Words</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Recognition Accuracy (%)</td>
<td>100.00</td>
<td>97.33</td>
<td>94.67</td>
<td>96.00</td>
<td>88.00</td>
<td>94.67</td>
<td>86.67</td>
</tr>
</tbody>
</table>

Fig 2: Sample data set

4.1 Results

In this case each PSL word was pre-processed individually and then presented to the network in an image matrix form. The program converts all binary matrices to bipolar form. After pre-processing, important information from input vectors are extracted as feature extraction. The important features are then fed to the classification engine. The table 1 describes the recognition rates achieved from each individual PSL word. The word wise recognition accuracy is depicted in figure 3.
4.2 Observations
The system recognized correctly most of the test patterns very similar to the target patterns. However, the system recognized the input characters having a significant level of noise (created manually).

In this implementation 94% recognition accuracy rates have been achieved. Findings of this implementation are summarized in tables 1 and 2.

5. CONCLUSION
The proposed shorthand recognition system is capable of recognizing specific set of shorthand words in general domain with 94% accuracy with hierarchical centroid and Neural
Network classifier combination. The experimental results show that the system can be used as a "working system" for shorthand recognition in this domain. We also experimented the recognition process with various popular feature extraction methods and classification techniques, but the results of these experiments, in all possible combinations of feature extraction methods and classification techniques, were below our expectations.

FUTURE WORK
The Shorthand recognizer system cannot be considered as a complete system, as for complete recognition of Shorthand language, we have to include more words and sentences available in Shorthand. These can be included in future. Since the limited data set that we have used, so in future, the total number of words can be extended to achieve good percentage of accuracy. There is scope for online recognition, where in recognition of the start and end points of the stroke become easy. The stenographer can use tablet and pen for taking the dictation leading to online English text generation. In general, stenographers take down dictation neatly because later generation of the English text depends on the readability of handwritten strokes. The work has global application in the areas of secret communication, legal transcription, compression etc.

6. ACKNOWLEDGMENTS
The authors thank to the referees for their valuable comments. They are also thankful to all those persons who have provided us their hand writings for this research.

7. REFERENCES


