

A Study of Handwritten Numerals for Profile Based Classification

Abstract:- Feature extraction is an important step of pattern recognition. This study uses profile based feature extraction and uses Simple Profile (with cropping and without cropping image samples), Contour based feature extraction for recognizing handwritten numerals. The features are computed by using 48 X 48 as a feature length classifier used Back propagation neural network for the classification. The classifier were trained and tested by using the MNIST Handwritten Numeral database. The average recognition rate of proposed system is observed as 91.10 %.

1. Introduction

Recognition of Handwriting number is an important area of research because of it is useful in many applications like reading machine for physically challenged person, postal code recognition, Banking system, Automatic text entry into computer for library automation, Automatic sorting of postal mail, bank cheques and other documents, Language processing. A recognition system must be writer independent, a recognition system can handle all possible variation in size shape, orientation, a recognition system may give low error rate and should have low rejection rate, the system can operate at high speed for commercial applications.

For the feature extraction and classification of handwritten number, large number of techniques existing but they are having mostly different strategy. Proposed study uses structural feature and neural network based architecture for recognition of handwritten numerals.

There are many feature extraction and classification scheme has been reported in literature survey. They mostly differ in feature extraction and classification schemes. Govindan et al. [1]. Feature extraction used for recognition include structural features, mathematical moments etc. classification scheme used include nearest neighbor schemes and feed forward networks. To make system more robust against various shapes researchers have used deformable models, multiple algorithms and learning. A recent survey of techniques is provided by plamondon & Srihari 2000[2]. Lam & Suen (1986) [3] have used a fast structural classifier and a relaxation-based scheme which uses deformation for matching. Knowledge based system using multiple expert has been used by Mai & Suen [4]. Kimura & Sridhar [5] developed a statistical classification technique and had used profiles and histograms of the direction vectors derived from the contours. Chen & Lieh [6] proposed a two-layer random graph-based scheme which used components and strokes as primitives. Jain & Zongkar [7] have proposed a recognition scheme using deformable templates. LeCun et al [8]. Suggested a novel

Backpropagation-based neural network architecture for handwritten zip code recognition. Knerr *et al* [9] suggested the use of neural network classifiers with single-layer training for recognition of handwritten numerals. Wang & Jean [10] suggested use of neural networks for resolving confusion between similar looking characters.

2. Process Overview

2.1 Feature Extraction

In this paper we have explore the problem related to stylic variation, similarity between numbers and style invariant feature. We present a study for recognition of handwritten numerals.

For this we have taken MNIST database which is consisting the 70000 image sample. There are 60000 samples for training purpose and 10000 for testing purpose.

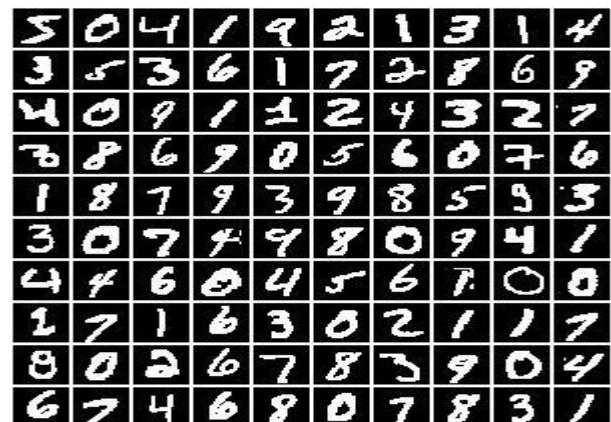


Figure 1: Original Samples of MNIST database

2.2 Feature Extraction

Shape of individual numerals can be characterized by considering distribution of dark pixels over image regions. Relative densities of the pixels over these regions characterize the shape of the character at reduced resolution.

Feature extraction method is used for taking the feature from the raw data which is available in the database and it is also used for finding the variability of the sample. A set of feature is help to severally extract it from another extracted.

For this purpose we used the following feature extraction method:

- a. Profile Based
 - i. Simple Profile
 - With Cropping
 - Without Cropping
 - ii Contour Based
- b. TAR Based

2.3 Profile based feature extraction

This method mainly counts the distance of pixel between the reverberate of the image and also find the edge of the number. For this scheme less memory are required. It describes the external shape of the number. Profile may be found by the two methods with cropping and without cropping.

Algorithm: For the Profile based

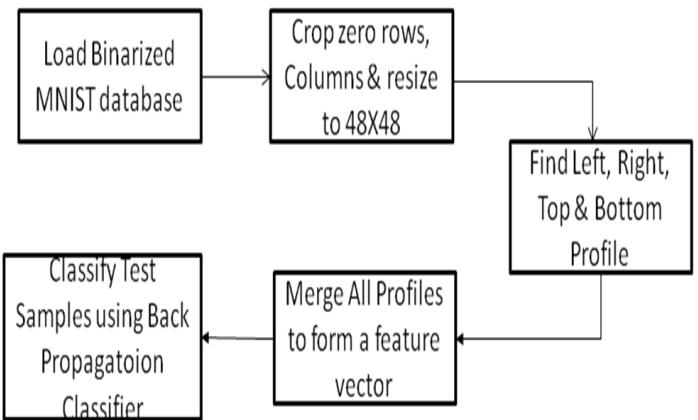


Figure 2: Block Diagram Profile based with Cropping

Firstly we load the MNIST database and crop the image and resize the image into 48*48 for best resolution than find the all profile of the image than merge for finding the feature vector of the image.

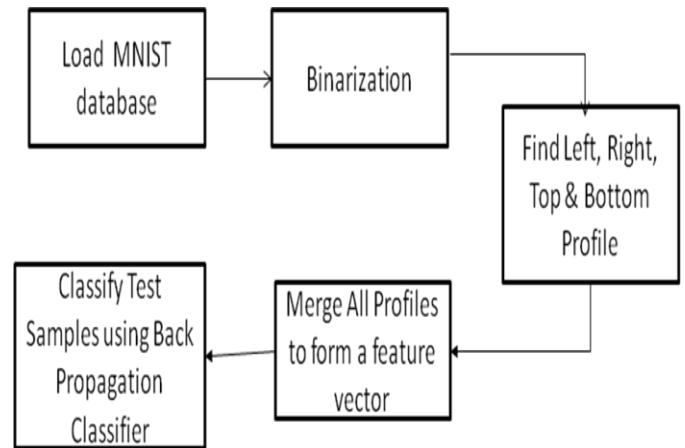


Figure 3: Block Diagram Profile based without cropping

For the Purpose of the Finding the profile without cropping the image we load the MNIST database than binarized the image than find the all profile of the image than merge all profile for the finding the feature vector of the image.



Figure 4 : Diagram for the Profile of Number 5

Algorithm: Contour based using Centre of Gravity

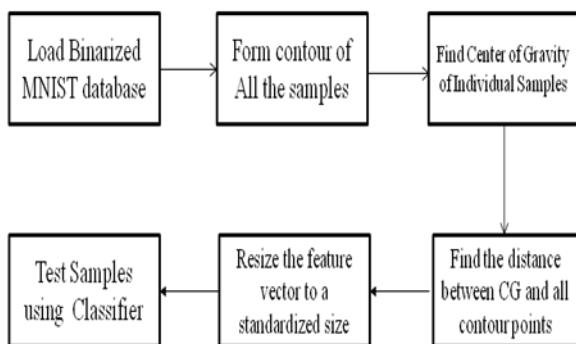


Figure 5: Block Diagram of Contour Based Feature extraction

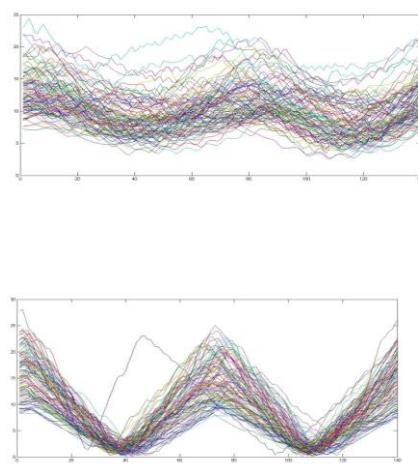


Figure 6: Distance Vector For Zero and One

For the finding of the feature of the image sample first load the MNIST data base than find the outer contour of the entire image sample than find the centre of gravity of the all individual samples. After finding this we find the distance between CG and all contour points than we resize the feature vector of a standard size than test the sample.

The search for the maximum and the minimum begins at the centre and extends to a fraction of the length of the component on the either side of the centre. a straight line is drawn between the two points and two vertical line are drawn connecting the minimum point on the top profile and the maximum point on the bottom profile to the respective top and bottom edges of the image.

Contour Image

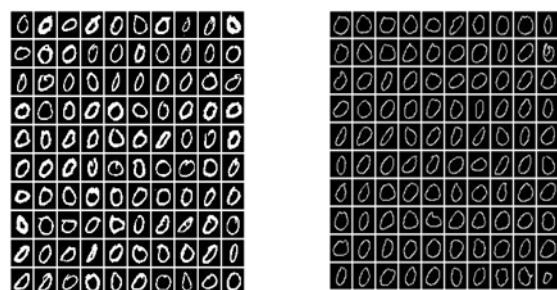


Figure 6: Formation of Outer Contour of the Image sample '0'

Algorithm TAR based feature extraction method

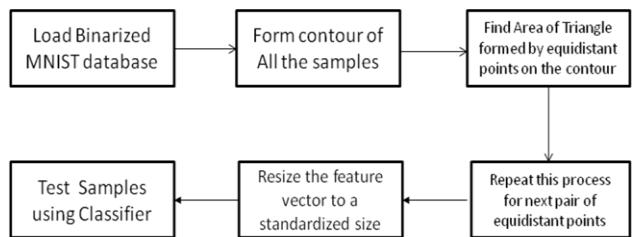


Figure 8: Block Diagram of TAR based feature extraction

Firstly we load the MNIST database and get the outer pixels or outer contour than find the centre of gravity of the image. After that we calculate distance of each pixel from CG than form feature vector using Distance vector algorithm than apply the classifier.

A neural network was over a large set of training patterns to cover a broad spectrum of variations in the numeral patterns. In the numeral patterns, neural network is required the efficient process in the training and testing. Each type of pen characters of the a different thickness. In order to reduce these variations, the thickness of the input pattern is first reduced and then find out the TAR of the image sample.

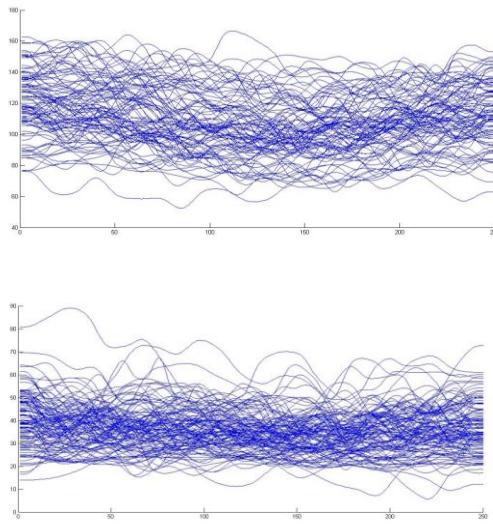


Figure 9: Feature Vector of Zero and One Using TAR

3. Normalization

Normalization is a form of size reduction process by which each of the numerals of varying sizes is cast into a form suitable for the recognizer. In our case we resize the image 28 x 28 pixels. Since the recognizer in ojut system is a LDA classifier neural net, the numerals must cast into a standard format. Each of the numeral image reduce the size of image.

4. Results

The system described above has been implemented in MATLAB. The neural has been used 60000 image sample for the training purpose and 10000 image sample for the testing purpose. The neural network was trained until the error become minimum. After the taken of confusion matrix we compare the all result we found the feature vector by TAR the result show the minimum. and we found the best result in the profile based algorithm.

The above numbers should be compared with the previous studies and results, obtained by others researchers, keeping complexity of the different target application in view.

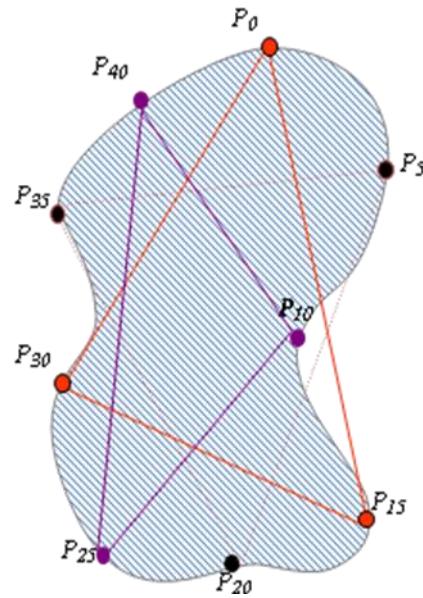


Figure 10. Formation of TAR

	Simple Profile With Cropping	Simple Profile Without cropping	Contour Based	TAR Based
BP	88.48%	91.30%	81.10%	57.79%

Table 3. Compression of all feature vector and classifier

5. Conclusion

We designed and implemented a profile based feature extraction and linear discriminate analysis classifier for handwritten numeral recognition. The study approach concerned with the problem of recognition of unconstraint, isolated handwritten numerals. The novel feature of this work is the approach followed for identification and integration of style specific information in the recognition scheme. Use of multiple classifiers is another significant feature of this work. Complete recognition architecture has been suggested in this work. This architecture is applicable for any character recognition problem. Experimental results show that this approach is potentially powerful.

Reference:

1. Govindan V K, Shivaprasad A P 1990 Character recognition – a review. *Pattern Recogn.* 23: 671–683
2. Hampshire JBII, Waibel A 1992 The meta-pi network: Building distributed knowledge representations for robust multisource pattern recognition. *IEEE Trans. Pattern Anal. Machine Intell.* PAMI-14: 751–769
3. Plamondon R, Srihari S N 2000 On-line and off-line handwriting recognition: comprehensive survey. *IEEE Trans. Pattern Anal. Machine Intell.* PAMI-22: 63–84
4. Lam L, Suen C Y 1986 Structural classification and relaxation matching of totally unconstrained handwritten ZIP codes. *Pattern Recogn.* 19: 15–19
5. Mai T, Suen C Y 1990 A generalised knowledge-based system for recognition of unconstrained hand-written numerals. *IEEE Trans. Syst., Man Cybern.* SMC-20: 835–848
6. Kimura F, Shridhar M 1991 Handwritten numerical recognition based on multiple algorithms. *Pattern Recogn.* 24: 969–983
7. Chen L-H, Lieh J R 1990 Handwritten character recognition using a two layer random graph model by relaxation matching. *Pattern Recogn.* 23: 1189–1205
8. Jain A K, Zongkar D 1997 Representation and recognition of handwritten digits using deformable templates. *IEEE Trans. Pattern Anal. Machine Intell.* PAMI-19: 1386–1391
9. LeCun Y, Boser B, Denker J S, Henderson D, Howard R B, Hubbard W, Jackel L D 1989 Backpropagation applied to Handwritten zip code recognition. *Neural Comput.* 1: 541–551
10. Knerr S, Personnaz L, Dreyfus G 1992 Handwritten digit recognition by neural networks with singlелayer training. *IEEE Trans. Neural Networks* 3: 303–314

Author Profile:

1. Akhilesh Pandey is an Asst. Professor in department of computer science and engineering Suresh Gyan Vihar University, Jaipur. He did his MCA from IGNOU in 2002 and after that he worked as a faculty member in different engineering college. After that he acquired his M. Tech. (CSE) at Sharda University, Gr. Noida, India.. His area of Interest is Pattern Recognition and neural network.
2. Rahul Yadav is a student in M.Tech. (EE) from Suresh Gyan Vihar University, Jaipur. He has a good interest in pattern recognition and artificial intelligence.