**COMPARISION OF ALGORITHMS FOR HANDWRITTEN DIGIT RECOGNITION**

**ABSTRACT:**

At present scenario, there is growing demand for a software system to recognize characters in a computer system when information is scanned through paper documents. This concentrates on adaptive methods that operate directly on size-normalized images and compares the relative merits of Neural Network based handwritten digit recognition system and Optical Character Recognizer.

**INTRODUCTION**

OCR (Optical Character Recognition) translates images of typewritten or handwritten characters into machine editable format. OCR reads damaged or low-quality codes and returns the best guess at what the code is. It is widely used as a form of information entry from printed paper data records, whether passport documents, invoices, bank statements, computerized receipts, business cards, mail, printouts of static data, or any suitable documentation. OCR does not deal with quality and sharpness of characters.

Handwriting Recognition is classified into two types: Off-line Recognition, deals with bitmap images from a camera and recognize the characters from the picture. Online Recognition, deals with x-y coordinates and try to recognize characters that are written on the smartphone. In off-line systems, the neural networks have been successfully used to achieve comparably high recognition accuracy levels.

The drawback of the entire system lies in the fact that results are highly dependent on font type of both the trained and tested characters.

The problem of pattern recognition can be solved by using machine learning. This gives computers the ability to learn without being explicitly programmed. We use supervised learning where a dataset is given and the correct output should look like as it is already known. We take a large number of handwritten digits, known as training examples and then develop a system which can learn from those training examples to automatically infer rules for recognition.

**METHODOLOGY**

High clarity image capture through TESSERACT-OCR technique. Accuracy can be increased if the output is constrained by a lexicon (a list of words that are allowed to occur in a document). Uses its dictionary to influence the character segmentation step.

Steps needed to detect handwritten digits: Create database of handwritten digits, for each digit extract HOG features and train a linear SVM and use the classifier trained to predict digits.

Digitize analog document using an optical scanner. Regions containing text are located and each symbol is extracted through segmentation. These are pre-processed to eliminate noise to facilitate feature extraction. Identity of each symbol is found by comparing extracted features with descriptions of symbol classes. Contextual information is used to reconstruct words of the original text.

The accuracy with which a neural network classifies patterns depends on how well it is trained. Training is a recursive process in which the connection weight between the neural networks is adjusted. The document is initially pre-processed so that it is noise free.

Pre-processing is used to discard irrelevant information in the input data, to increase the speed and accuracy. It consists of binarization, normalization, sampling, smoothing and denoising. Text localization is done using region extractor. Features are extracted from the localized regions. Apply classification algorithms to recognize the input.

**APPLICATIONS**

1. **Institutional repositories** are digital collections of the outputs created within a university or research institution. It is an online locale of intellectual data of an institution, especially a research institution where it is collected, preserved and aired. It helps to open up the outputs of an institution and give it visibility and more impact on worldwide level.

2. **Invoice imaging** is widely used in many businesses applications to keep track of financial records and prevent a backlog of payments from pilling up. In government agencies and independent organizations, OCR simplifies data collection and analysis, among other processes.

3.**Automatic number plate recognition** [6] is used as a mass surveillance technique making use of optical character recognition on images to identify vehicle registration plates. ANPR has also been made to store the images captured by the cameras including the numbers captured from the license plate.

4. The **legal industry** is also one of the beneficiaries of the OCR technology. OCR is used to digitize documents and directly entered into a computer database.

5. Another important application of OCR is in **banking**, where it is used to process cheques without human involvement. Cheque can be inserted into a machine where the system scans the amount to be issued and the correct amount of accessed as necessary

6. **Healthcare** has also seen an increase in the use of OCR technology to process paperwork. Healthcare professionals always have to deal with large volumes of forms for each patient, including insurance forms as well as general health forms. To keep up with all of this information, it is useful to input relevant data into an electronic database that can be accessed as necessary.

**CONCLUSION**

This concentrates on adaptive methods that operate directly on size-normalized images. Compares the relative merits of Neural Network based handwritten digit recognition system and Optical Character Recognizer. The problems faced are due to diverse text patterns and various background interferences and the image distortion. The printed documents and handwritten documents are transformed to ASCII files for the purpose of compact storage, editing, fast retrieval and other file manipulations through the use of neural networks and optical character recognition.

**REFERENCES**

[1] A. S. Sawant, “Script Independent Text Pre-processing and Segmentation for OCR,” Int. Conf. Electr. Electron. Signals, Commun. Optim. - 2015, pp. 1–5, 2015.

[2] V. Kieu, F. Cloppet, and N. Vincent, “OCR Accuracy Prediction Method Based on Blur Estimation,” 2016 12th IAPR Work. Doc. Anal. Syst., pp. 317–322, 2016.

[3] J. B. Pedersen, K. Nasrollahi, and T. B. Moeslund, “Quality Inspection of Printed Texts,” IWSSP 2016- 23rd Int. Conf. Syst. Image Process. 23-25 May 2016, Bratislava, Slovakia, pp. 6–9, 2016.

[4] A. F. Mollah, N. Majumder, S. Basu, and M. Nasipuri, “Design of an Optical Character Recognition System for Camera- based Handheld Devices,” IJCSI, vol. 8, no. 4, pp. 283–289, 2011.

[5] B. Jain and M. Borah, “A Comparison Paper on Skew Detection of Scanned Document Images Based on Horizontal and Vertical,” IJSRP, vol. 4, no. 6, pp. 4–7, 2014.

[6] E. N. Bhatia, “Optical Character Recognition Techniques : A Review,” IJARCSSE, vol. 4, no. 5, pp. 1219–1223, 2014.