

CHARACTER RECOGNITION SYSTEM – A FRAMEWORK OF IMPLEMENTATION**A.A. Tayade¹ and R.J. Ramteke²**¹G.S. Science, Arts and Commerce College Khamgaon, MS, India²School of Computer Science, K.B.C. North Maharashtra University, Jalgaon, MS, India¹arvindtayade40@gmail.com, ²rakeshramteke@yahoo.co.in**ABSTRACT**

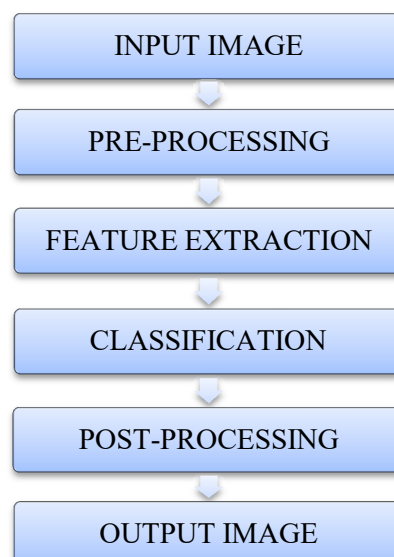
Character acknowledgment framework is significant part of the of example acknowledgment. Character acknowledgment is a stepwise cycle of preprocessing, highlight extraction and characterization. In character acknowledgment, precision diminished because of certain limitations like mathematical misalignment through character style varieties, clamor, and undesirable data in picture, size of the picture and in particular the fluctuating picture foundation. Expanding the exactness and improving the FRT model, need to choice of the legitimate element extraction system and appropriate classifiers. In this research, we focus on the element extraction calculations and utilized number of classifier to assess the framework. For improving the acknowledgment precision, lessening the component vector length, decreasing the execution time, we proposed not many calculations or philosophy for character acknowledgment.

Keywords: feature extraction, preprocessing, classification

Introduction

In this research system, the character acknowledgment calculation incorporates LBP, directional component, Randon, Lenna dhamma are presented. The system will talk about here have ability of picture commotion, brightening, impact of scaling and revolution invariant. The main significant advance in any character acknowledgment framework is pre-handling followed by division and highlight extraction. Pre-handling includes with the means that are needed to shape the info picture into a structure reasonable for division. In division step, the information picture is divided into singular characters and afterward, each

character is resized into $m \times n$ pixels towards preparing the organization. The choice of proper component extraction technique is likely the absolute most significant factor in accomplishing high acknowledgment execution. A few strategies for highlight extraction for character acknowledgment have been accounted for in the writing. The overall Square graph for character highlights recognition fundamentally comprise of some essential squares like information picture, pre-preparing, include extraction, grouping, post-handling lastly get yield picture. The overall strategy of character acknowledgment is as demonstrated beneath:



Block diagram of the proposed character recognition system

Local Binary Patterns

The element vector would now be able to be prepared utilizing the Help vector machine, outrageous learning machines, or some other AI calculation to group pictures. Such classifiers can be utilized for character acknowledgment or texture examination.

The nearby twofold example administrator is a picture administrator, which changes a picture into an exhibit or picture of number marks portraying limited scope appearance of the picture. It has demonstrated to be profoundly discriminative and its central issues of revenue, specifically its invariance to monotonic dark level changes and computational capability, make it reasonable for requesting picture investigation errands. The fundamental nearby parallel example administrator, was dependent with the understanding that surface has locally two reciprocal angles, an example and its solidarity. LBP highlight extraction comprises of two chief advances: the LBP change, and the pooling of LBP into histogram portrayal of a picture. As clarified in [3] dark scale invariance is accomplished due to the distinction of the power of the adjoining pixel to that of the focal pixel. It additionally embodies the nearby math at every pixel by encoding binarized contrasts with pixels of its neighborhood area:

$$LBP(P, R, t) = \sum_{p=0}^{p-1} S_t(g_p - g_c) 2^p$$

Where g_c is the focal pixel being encoded, g_p are P evenly and consistently examined focuses on the fringe of the roundabout space of sweep R around g_c , and st is a binarization work boundary by t . The examining of g_p is performed with bilinear insertion. t , which in the standard definition is viewed as nothing, is a boundary that decides when neighborhood contrasts are viewed as large enough for thought.

In our LBP, the first form of the nearby double example administrator works in a 3×3 pixel square of a picture. The pixels in this square are edge by its middle pixel esteem, duplicated by forces of two and afterward added to get a mark for the middle pixel. As the local comprises of 8 pixels, a sum of $2^8 = 256$ distinct marks can be acquired relying upon the

general dark estimations of the middle and the pixels around there.

The LBP highlight vector, in its least difficult structure, is made in the accompanying way:

- Divide the analyzed window into cells.
- For every pixel in a cell, contrast the pixel with every one of its 8 neighbors (to its left side top, left-center, left-base, right-top, and so on) Follow the pixels along a circle, for example clockwise or counter-clockwise.
- Where the middle pixel's worth is more noteworthy than the neighbor's worth, state "0". Something else, state "1". This gives a 8-digit double number (which is generally changed over to decimal for comfort).
- Compute the histogram, over the phone, of the recurrence of each "number" happening (i.e., every blend of which pixels are more modest and which are more prominent than the middle). This histogram can be viewed as a 256-dimensional element vector.
- Optionally standardize the histogram.
- Concatenate (standardized) histograms, all things considered. This gives a component vector for the whole window.

Radon Transformation

As of late, the Radon transform have gotten a lot of consideration. This transform can change two dimensional pictures with lines into a space of conceivable line boundaries, where each line in the picture will give a pinnacle situated at the relating line boundaries. These have lead to many line discovery applications inside picture handling, PC vision, and seismic. The Radon Change is a basic instrument which is utilized in different applications, for example, radar imaging, geophysical imaging, nondestructive testing and clinical imaging.

Discrete Cosine Transform (DCT)

DCT is the most broadly utilized transform in the picture preparing applications for include extraction. DCT addresses an amount of sinusoids of fluctuating sizes and frequencies. The methodology includes taking the change of a picture in general and isolating the applicable coefficients. The image is partitioned into a 16×16 sliding window where each 8×8 square is exposed to discrete cosine transform (DCT)

to ascertain the recurrence segments. The DCT of a picture essentially comprises of three recurrence segments specifically low, center, high each containing some detail and data in a picture. The higher recurrence parts offer lesser to the picture subsequently their qualities can be diminished to zero by quantization.

The DCT equations can be represented mathematically as,

$$y(k) = w(k) \sum_{n=1}^N x(n) \cos \left[\frac{\pi(2n-1)k-1}{2N} \right],$$

$$k = 1, 2, \dots, N$$

where x,y is a matrix of row or column, N is the size of the matrix.

The input image for the DCT algorithm is a color image. Using the standard formula, it is transformed to a grayscale image.

$$I = 0.299R + 0.587G + 0.114B.$$

Experimental Analysis of Character Recognition Technique

Two experimental setup has been presented here as shown below

Experimental setup 1: LBP (local Binary Pattern)

The LBP for feature extraction. it is theoretically very simple yet efficient approach for gray scale and rotation invariant texture classification and nonparametric discrimination of sample and prototype distributions. To performing the result take 59 features for classification and various (KNN, SVM AND ITS VARIANTS) classifier use to classify the data. We shows the results in four category i.e. numeric dataset, midbar characters, end bar characters and no bar characters. First upon we display data set wise results then shows the comparative analysis of various dataset results analysis.

Experimental setup 2: Lanna dhamma

Lanna Dhamma alphabet was used mainly for religious communication in the ancient Lanna Kingdom of Thailand. The old manuscripts using this alphabet are gradually decayed. It is desirable to preserve these valuable manuscripts in machine-encoded text files. Existing works used optical character

recognition (OCR) methods based on wavelet transform for recognition of handwritten Lanna Dhamma characters. This method contains Tchebichef, Hahn, DCT transform we are use these three combinations for features extraction use by proposed technique. Charterer dived into three sections mid bar, no bar and end bar.

Experimental setup 3

In this, set up used the feature vector extracting from sum of image columns and sum of image row. The row wise sum calculates and treat as the feature vector using this feature vector table labile this table for predictions. The calculation result displays the various variang result and perform the good approach. This set up also divided in to three character category i.e. mid bar,no bar and end bar character. Numeric data also calculated with different font size and style. For classification various classifier are used. SVM, KNN etc. following table shows the results of experimental setup 3. In this approach we use the only sum of character image the regionprop() function use for finding the region from image. Find the region of character and resize it 20 X 20 pixel image and perform the column wise addition this addition treated as the feature vector and achieves the following recognition accuracy which is greater than above methods with less feature vector i.e. 20.

Experimental setup 4: Feature extraction from specific region of text

In LBP based method, the LBP features from whole image is used, but in such a case take the features from the unwanted area. If character is small then the then unnecessary region also computed. To solve the such types of problem, there is a need to find specific region from whole character image.

Conclusion

In this research, the various methods are used to recognition of character. Some of the methods are more attractive as depending the database used for extraction. Here, the data validation technique, average recognition accuracy as been calculated based on the techniques such as LBP, Lanna dhama. Here, the LBP method is used in two-way approach. In first approach, we directly implement the

LBP on image and in another approach on it, then accuracy increases respectively. important information cropped and apply LBP

References

1. Bansal, R.S. (1995), Devnagari document processing, IEEE Explore.
2. Gian Marco D.D., (1998). An Analog VLSI Implementation of a feature extraction for real time OCR, IEEE, vol. 33.
3. Chan W.C., (1998). Off-Line Handwritten Chinese Character Recognition as a Compound Bayes Decision Problem, IEEE, vol. 20.
4. Xu Y.Y., (1998). Multilingual Handwritten Character Recognition by Bayesian Decision-Based Neural Networks, IEEE, vol. 46.
5. Hagita N. (1998). Text-Line Extraction and Character Recognition of Document Headlines With Graphical Designs Using Complementary Similarity Measure, IEEE, vol. 20.
6. Chin R.T. (1998). A Bayesian Framework for Deformable Pattern Recognition with Application to Handwritten Character Recognition, vol. 20.
7. Yuichi Nakatani H. M. (1999). Online Recognition of Handwritten Hiragana Characters Based Upon a Complex Autoregressive Model, vol. 21.
8. Nei Kato H.A. (1999). A Handwritten Character Recognition System Using Directional Element Feature and Asymmetric Mahalanobis Distance, IEEE, vol. 21.
9. Dlay S.S. (1999). Analog Design of a New Neural Network for Optical Character Recognition, IEEE, vol. 10.
10. Nakagawa M. (2000). Precise Candidate Selection for Large character set Recognition, IEEE, vol. 22.
11. Rahman F.R. (2001). A Multiexpert Framework for Character Recognition: A Novel Application of Clifford Networks, IEEE, vol. 12.
12. Toru Wakahara A.T. (2001). Affine-Invariant Recognition of Gray-Scale Characters Using Global Affine Transformation Correlation, IEEE, vol. 23.
13. D. Doermann (2004). Machine Printed Text and Handwriting Identification in Noisy Document Images, IEEE, vol. 28.
14. Roy P.P. (2004). Multioriented and Curved Text Lines Extraction From Indian Documents, IEEE, vol. 34.