iJRASET
International Journal For Research in Applied Science and Engineering Technology

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 6 **Issue:** IV **Month of publication:** April 2018

**DOI:** http://doi.org/10.22214/ijraset.2018.4652

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887*
*Volume 6 Issue IV, April 2018- Available at www.ijraset.com*

# A Survey on Evaluating Handwritten Iterative Mathematical Expressions

Tina Vaz[1] , Nagaraj Vernekar[2]
[1]*Goa College of Engineering*
[2]*Associate Professor Goa College of Engineering*

***Abstract: Handwritten Mathematical Expression Recognition have become one of the most challenging, fascinating and growing research area in the field of Pattern Recognition. This paper focuses on the work carried out by several researchers in the task of recognizing Handwritten Mathematical Expressions. An attempt is made to analyze different stages of recognition process. In particular, this work tries to concentrate on various feature extraction methods and the recognition system and a comparative analysis is been prepared. This research is carried out to perform recognition of Handwritten Iterative Mathematical Expressions and get the computational results.***
***Keywords: Handwritten Mathematical Expression; Feature extraction; SVM.***

## I. INTRODUCTION

Pattern recognition is concerned with finding of uniformities in data through use of computer algorithms and using it to take certain actions such as data classification. For handwritten mathematical expression recognition the goal is to build a machine that identifies different symbols in the expression. But the problems arises in different variations in writing. The best solution can be obtained by adopting Machine learning approaches [1].

Handwritten mathematical expression recognition has received inordinate attention in today's research area due to importance of mathematics in many scientific fields which demands for a mechanism for converting handwritten mathematical expressions into an electronic format for faster processing. The researchers have proposed many tools towards mathematical symbol and expression recognition. Tools, for example, Latex and MathML requires knowledge of predefined sets of keywords [2][3].

The recognition of handwritten mathematical expression is a difficult and challenging task due to various reasons such as large character set, variations in style and size of symbols by different writers, twisted semantics and two dimensional layout of symbols. There is no availability of public dataset of handwritten ME's. Therefore researchers try to gather their own dataset which might limit the problem to a certain domain.

The recognition can be online or offline. Online recognition works on ink stokes obtained directly from pen devices whereas offline recognition works on image containing expressions. Recognizing an expression from an image is more challenging [4] due to the fact that online information provides complete details the way an expression is written.

Our research focuses on analyzing different approaches used towards offline handwritten mathematical symbol and expression recognition.

## II. RECOGNITION OF OFFLINE HANDWRITTEN MATHEMATICAL EXPRESSIONS

This section deals with the work carried out by various researchers towards the task of recognizing handwritten mathematical expressions. The focus is made on different techniques and methods used in different phases of recognition.

R.Padmapriya and S. Karpagavalli [5] proposed HME recognition system carried out in different phases which include data collection, preprocessing, segmentation, feature extraction, symbol classification and also mathematical expression recognition. A data set of 50 simple algebraic expressions written by 10 writers, each equation with 10 to 15 symbols converting 23 unique symbols are collected. Preprocessing involves binarization, noise reduction, size-normalization and skew detection and correction. In segmentation process connected components analysis technique is adapted. The features set consists of zonal, structural, skeleton based and directional features. The recognition process is carried with the help of Support Vector Machine (SVM) and Multilayer Perceptron (MLP) classifiers. The recognition rate achieved through MLP classifier is 92% and SVM classification is 85%.

Manisha Bharambe [6] focuses on recognition of handwritten logical mathematical expressions. The work deals with two dimensional structure of expression having subscripts. Data is collected from 20 writers from different fields, each symbol written by each writer 10 times. Database of 2000 symbols and 1560 letters have been collected from different writers. The datasheets are scanned and individual symbol images are cropped from scanned image manually, resulting in gray scale image of symbol.

Preprocessing stage involves filtering, morphological operations and image thinning. Bounding box is used for segmenting each character from the expression. The features set consists of normalized chain code, moment invariant features, density feature and projection histogram. Recognition is accomplished using SVM classifier. The linear kernel bsvm2 is used for training purpose and a model is built that assigns a class label for each category. A set of 50 handwritten logical expressions have been tested and overall average recognition rate of the ME recognition is 93.8%.

Gharde, et al. [7] have proposed offline HME recognition system for simple Mathematical Equations. The dataset consists of 237 symbols which are collected from 28 simple Mathematical expressions. This dataset is divided into dataset 1 and dataset 2 for training purpose.

Preprocessing involves smoothing, enhancing, filtering, cleaning-up of image. Bounding box method is used for segmentation purpose. A focus is made on evaluation of various methods used for feature extraction and recognition system. A projection histogram based technique is used for feature extraction.

For recognition purpose, SVM classifier is used. Using projection profile and support vector machine two different datasets are recognized. The recognition rate achieved is 97.58% and 98.40%.

Sagar Shinde, Rajendra Waghulade [8] has made an effort to implement feed-forward back propagation neural network for the recognition of handwritten mathematical expressions.

The system presents an approach to recognize handwritten straight line equation and quadratic equation with the focus on improving recognition rate, performance, processing time and accuracy.

Preprocessing is carried out on the scanned input image that includes noise removal, binarization, skew correction, contrast enhancement and regions identification containing the mathematical expression. Line, word and character segmentation is used for image decomposition into individual characters.

Centroid and bounding box features are extracted for each character. The training of neural network is done using gradient descent with momentum and adaptive learning. Performance of developed system is been measured and systems accuracy and learning rate is been calculated.

Nicolas D. Jimenez, Lan Nguyen[9] has made an effort to show that pyramids of oriented gradients (PHOG) are effective features for Handwritten Mathematical symbols recognition. The best results are obtained along with one against one SVM classifier. The CHROHME dataset is used for training the classifier, which contains 22000 character samples and 1400 equations. The analysis are limited to 59 characters.

The generalization accuracy achieved on these characters is 96%. They demonstrated that the classifier gracefully generalize symbols and images of 75 handwritten symbols written by new users are also tested.

Iffath Fathima S, Ashoka K[10] has proposed an approach for mathematical symbols recognition. The experiment is carried out over handwritten as well as printed mathematical symbols.

The dataset includes math symbols as well as alphabets of English, Latin, Greek languages. An own dataset is used for handwritten symbols whereas standard Infty MDB-1 dataset for printed images. The work focuses on using character geometry as a feature extraction technique due to its support for unmistakable component extraction strategies.

The classifiers Support Vector Machines (SVM) and K-nearest neighbour (KNN) are used for classification. A comparative analysis is been carried out using both classifiers.

The accuracy of SVM and KNN classifiers over printed dataset is 93% and 94% respectively and handwritten dataset is 88% and 91%. The experimental results show that KNN has better accuracy as compared to SVM classifier. The work also stated that efficiency of KNN decreases as the dataset increases.

## III. A COMPARATIVE STUDY

In this section, we propose a comparative study of various feature extraction techniques and the classifiers used for classification. Among different phases carried towards the recognition process, feature extraction and classification phase may have great impact on overall accuracy and recognition rate of the system. Feature extraction is used to extract certain features of interest that will help the classifier in distinct symbol classification. Table 1. depicts a comparative study of various feature extraction techniques and classifiers gathered from research papers in the area of mathematical symbol and mathematical expressions recognition. The analysis suggests that projection histogram and pyramids of oriented gradients most suitable features that yields better accuracy and support vector machine an appropriate classification technique. SVM have gained importance in the field of pattern classification. They are supervised learning methods used for classification problem.

TABLE I. COMPARATIVE STUDY OF VARIOUS FEATURES AND CLASSIFIERS

| Author | Features set | Classifier | Recognition rate |
|---|---|---|---|
| R.Padmapriya,S. Karpagavalli | Zonal, Structural, Skeleton based and Directional features | SVM | 85% |
| | | MLP | 92% |
| Gharde, et al. | Projection histogram | SVM | 97.58% and 98.40% for two different datasets |
| Manisha Bharambe | Normalized chain code, Moment invariant features, Density feature, Projection histogram | SVM | 93.8% |
| Sagar Shinde, Rajendra Waghulade | Centroid and bounding box | Artificial Neural Network | - |
| Nicolas D. Jimenez, Lan Nguyen | Pyramids of oriented gradients(PHOG) | SVM | 96% |
| Iffath Fathima S, Ashoka K | Character Geometry | SVM | 88% |
| | | KNN | 91% |

## IV. CONCLUSION

In This paper, we have presented a survey in the direction of recognizing handwritten mathematical symbols and expressions. Different phases towards the recognition process are analyzed out of which feature extraction and classification phase is given importance due to its great impact on overall accuracy and recognition rate. Among various feature extraction techniques, projection histogram and pyramids of oriented gradients are found to be most suitable features that yields better accuracy and support vector machines an appropriate classification technique.

## REFERENCES

[1] Richard Zanibbi, Dorothea Blostein “Recognition and Retrieval of Mathematical Expressions” International Journal on Document Analysis and Recognition, Springer-Verlag, 2011.

[2] Ahmad Montaser Awal, Harold Mouchere, Christian Viard Gaudin “Towards Handwritten Mathematical Expression Recognition” 10th International Conference on Document Analysis and Recognition. IEEE 978- 07095, 2009.

[3] Ahmad Montaser Awal, Harold Mouchere, Christian Viard Gaudin “The Problem of Handwritten Mathematical Expression Recognition Evaluation”. ISBN, 978-0- 7695-4221-8,2010.

[4] B. Keshari, S. Watt, “Hybrid Mathematical Symbol Recognition using Support Vector Machine”, Analysis and Recognition – volume 02 Pages859-863, 2007.

[5] R.Padmapriya, S. Karpagavalli “Offline Handwritten Mathematical Expression Recognition” International Journal of Innovative Research in Computer and Communication Engineering, Vol. 4, Issue 1, January 2016.

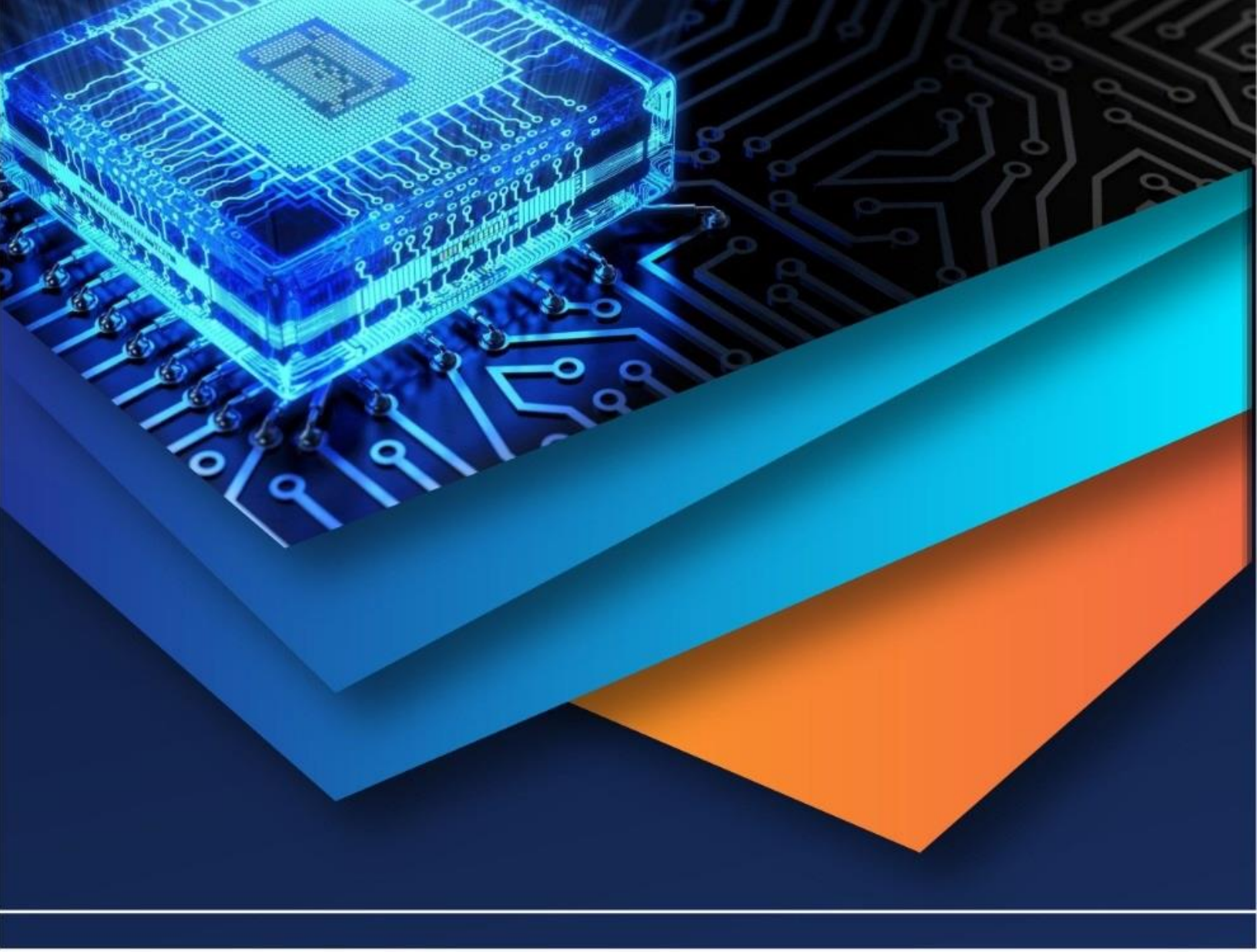
[6] Manisha Bharambe “Recognition of Offline Handwritten Mathematical Expressions” International Journal of Computer Applications (0975 – 8887) , National conference on Digital Image and Signal Processing, DISP 2015.

[7] Gharde, S. S., Baviskar, P. V., & Adhiya, K. P. “Identification of Handwritten Simple Mathematical Equation Based on SVM and Projection Histogram” International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-3, Issue-2, May 2013.

[8] Sagar Shinde, Rajendra Waghulade “Handwritten Mathematical Expressions Recognition using Back Propagation Artificial Neural Network” Volume 4– No.7, March 2016 Communications on Applied Electronics (CAE).

[9] Nicolas D. Jimenez and Lan Nguyen “Recognition of Handwritten Mathematical Symbols with PHOG Features”, http://cs229.stanford.edu.

[10] Iffath Fathima S, Ashoka K “Machine Learning Approach for Recognition of Mathematical Symbols” International Journal of Scientific Research Engineering & Technology (IJSRET), ISSN 2278 – 0882 Volume 6, Issue 8, August 2017.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY