Recognition Of Handwritten Mathematical Equations

Neelam Pal[1], Lavanya Agrawal[2], Prachi Gupta[3], Upasana Pandey[4]
[1, 2, 3, 4] Dept of Computer Science & Engineering
[1, 2, 3, 4] IMS Engineering College, Ghaziabad, Uttar Pradesh,India

Abstract- *One goal of the present work is to analyze and discover features that can best represent handwrittencharacters. We have decomposed each individual symbol into basic elements .Raw recognition may then be performed on the basic elements.The aim of this project is to develop an assistance system that analyses the handwritten mathematical equations based on handwriting recognition algorithms*

While large number of researches have been done for digits and characters recognition, much less progress has been made surrounding handwritten equation recognition. While typing is much faster than writing by hand, math equations have the opposite property: writing mathematical equations by hand is more efficient than typesetting them. Furthermore, whereas handwritten equations are human readable, the "code" of typesetting languages are often highly nested and difficult to edit. These problems of typesetting present an opportunity for improving the workflow of writing math equations digitally.

Handwritten character or symbol recognition is one of the application in pattern classification. It is generally easy for anyone to recognize handwritten or characters and symbols but it is difficult for a computer to recognize them. This difficulty can be overcome by adopting machine learning approach by designing a system that recognizes the patterns. Pattern classification involves features extraction, concept behind the observation and classifier. For this purpose, character geometry as feature extraction technique and two classifiers Support Vector Machines (SVM) and K-nearest neighbor (KNN) are used. Two classifiers are used for the comparative analysis.

Keywords- SVM, KNN, Machine learning

I. INTRODUCTION

1.1 IDENTIFYING THE PROBLEM

Handwriting recognition has been studied for about fifty years and has been used in hand-held devices post office scanners etc . With the recent development of electronic tablets, pen-tip movement can be captured more accurately, making possible the capture of not only x, y coordinates, but pen pressure as well.

Pen-based computing combined with handwriting recognition has become an important research topic and has attracted significant attention in the last few decades,being considered as a key development needed for the next generation of PDAs and tablet PCs. Mathematics inputting and editing have posed problems due to their two dimensional structure and large sets of symbols. These typically consist of special symbols and Greek letters in addition to English letters and digits. The commonly used keyboard input is thus insufficient for the input of such a large set of symbols, which has led to the desire for other input methods. The most widely used such system is LATEX. A set of keywords is defined in LATEX for the representation of special symbols and characters. Most scientific and engineering publications contain mathematical symbols and expressions. Recognition of handwritten mathematics would not only require less effort in writing technical documents but could also be used to transfer existing handwritten documents into electronic format and between machines

when needed. Therefore handwritten mathematics recognition is one of the key forces that drive the information transformation between human and machine and among machines.

1.2 PROBLEM STATEMENT

We examined the problem of machine recognition of handwritten mathematical symbols. We concentrated on the following areas: reducing the amount of computation,identifying discriminative features between symbols and building recognition models. One goal should be to reduce the amount of computation for recognition when a large number of symbols is used. Feature extraction can be an important step for recognition, since features distinguish the individual characters. So far, researchers have used a wide range of different features, such as geometry features, global features and so on. Main goal of the present work is to analyze and discover features that can best represent handwritten

characters. We have decomposed each individual symbol into basic elements.Raw recognition may then be performed on the basic elements.

II. IMPLEMENTATION OF THE PROPOSED SYSTEM

The system will go through five stages before prouducing the final result.The five stages are given below-

Input image: Input picture can be a picture comprising of either a character, image or expressions. The information picture ought to be in a .png augmentation necessarily.

Preprocessing of input image: Picture pre-dealing is compulsory for any photo based applications. The precision and consolidating rate of such frameworks must be on a very basic level high in order to ensure the achievement of the resulting steps. In any case, more often than not, the noteworthiness of these systems stay ignored which brings about second rate results. The objective of the pre-handling step is to construct necessary information.

Feature extraction: This is a sort of dimensionality decreasing that successfully addresses interesting parts of a photo as a little component vector. This approach is useful when picture sizes are huge and a reduced segment depiction is required to quickly complete assignments, for instance, picture organizing and recovery. Character Geometry Feature Extraction Technique is used for the component extraction since it is one which support the component extraction strategies which does not have any mistakes.

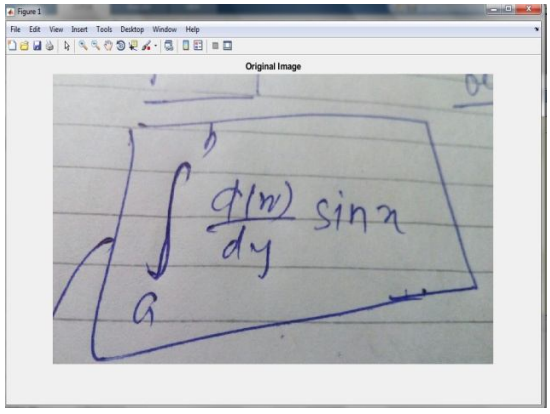
Performing SVM and KNN classification: KNN count is a flexible and clear estimation which arranges the given get ready cases in perspective of its neighbors. For collecting the new case this count figures the Euclidian division with the new representation and recognizes its neighbors after that it consigns the class in light of the k regard .k regard will be customer portrayed. This estimation reconstruct the class that addresses the most outrageous of the k cases. SVMs (Support Vector Machines) are an important technique for data classification. A classification task generally incorporates disengaging data into get ready and testing sets. Each event in the planning set contains one "target regard" (i.e. the class names) and a set of "attributes" (i.e. the components or watched factors). The goal of SVM is to explain a model (in light of the arrangement data) which predicts the target estimations of the test data given only the test data properties. SVM requires that each data case is addressed as a vector of real numbers. Scaling before applying SVM is tough. The rule of SVM depends upon a straight separation in a high estimation feature space where data are mapped to consider the possible non-linearity of the issue.

Printing the recognized symbol as Output: The expression, symbol or character given as the input picture is behold and imprinted in the command window. Recognized image as bestow is a consequence of the considerable number of

III. EXPERIMENTAL RESULTS

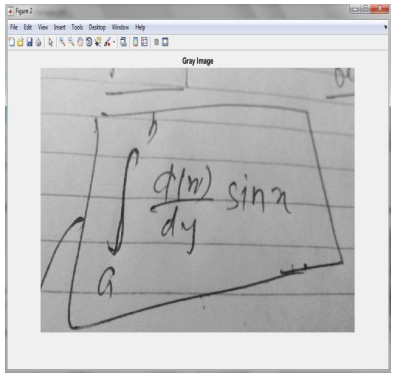
This project only aims to achieve part of the mathematical expression recognition. Indeed, we do not recognize the symbols but rather put them into classes, given the structure of the formula. On the other hand, we cannot expect a perfect recognition of the structure, provided that the symbols identities are not known. Therefore, we should not only count the mistakes of the system, but also estimate how confident it is on the expected recognition.

2.1 First Stage



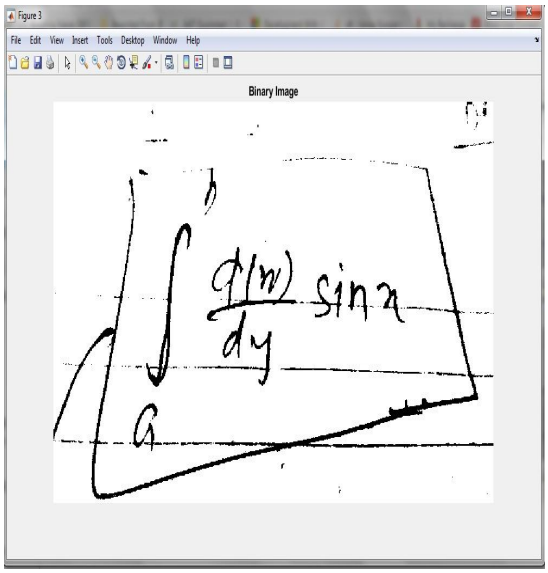
In the very first stage we have given the input to the system through a real time device called webcam.The webcam has captured the image and has presented in its raw form.

2.2 Second Stage



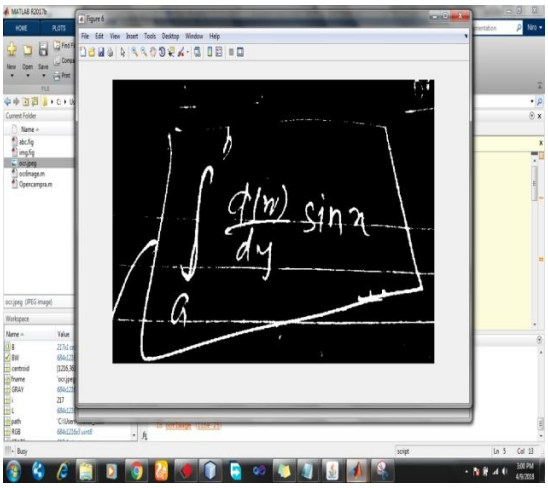
In the second stage after the processor has received the image in its raw form or its original form , it proceeds furtherto convert it in a greyscale image as shown in image. A greyscale image is one in which the value of each pixel is a single sample representing only an amount of light,that is,it carries only intensity information.

2. 3 Third Stage



After the image has been converted into a grey scale image, it is processed further to convert it in a binary image. A binary image is a digital image that has only two possible values for each pixel. Typically, the two colors used for binary image are
Black and white. It is one of the intermediate results of the project.

2.3 Final Stage



In the last stage image is processed after converting it into a binary image to produce the final results. The image above is the output of the original image we gave.It has been converted into the final product and the system has successfully recognized the handwritten mathematical symbols.

IV. CONCLUSION

Among different feature extraction techniques and strategies, Character geometry is selected as the feature extraction technique. Character geometry feature extraction technique is one which supports other different feature extraction techniques. SVM and KNN classifiers are used for the classification. After numerous executions it has been found that classifier has great correctness compared to SVM as classifier. The efficiency of KNN decreases with the increase in dataset.

V. ACKNOWLEDGMENT

We are earnestly grateful to our mentor **Dr. UPASANA PANDEY ma'am,** Associate Professor Department of computer science and engineering, IMS Engineering College for guiding us at each and every step to make this project a successful one. Finally we will like to express heartiest gratefulness to almighty and to our parents for supporting us.

REFERENCES

[1] Richard Zanibbi & Dorothea Blostein, "Recognition and retrieval of mathematical expressions", IJDAR, SpringerVerlag, 2011.
[2] Zhao Xuejun, Liu Xinyul, Zheng Shenglingl, Pan Baochang and Yuan Y.Tang, "On-line Recognition Handwritten Mat hematical Symbols", IEEE, 1997.
[3] Dorothea Blostein & Ann Grbavec, " Handbook on Optical Character Recognition and Document Image Analysis", (Chapter22)-"Recognition of Mathematical Notation", World Scientific Publishing Company, 1996
[4] J. S. Raikwal & Kanak Saxena ," Performance Evaluation of SVM and K-Nearest Neighbor Algorithm over Medical Data set " In International Journal of Computer Applications (0975 – 8887) Volume 50 – No.14, July 2012M
[5] Xiaofang Xie ,"On the Recognition of Handwritten Mathematical Models".