Asian Journal of Information Technology 15 (16): 2943-2947, 2016
ISSN: 1682-3915

Graphical RepreSentation of Handwritten Mathematical Expression Recognized from Offline Video Lectures

[1]R. Suganya and Dhanya [2]M. Dhanalakshmy
[1]Department of Computer Science and Engineering, Amrita School of Engineering, Coimbatore, India
[2]Amrita Vishwa Vidyapeetham, Amrita University, Amrita, India

Abstract: Mathematics is a universal language in terms of science and technology. It gives the exact information than any other languages. Hence recognition of mathematical expression is one of the hot areas of research in pattern recognition and machine language. In this study, recognition of handwritten mathematical expression with the corresponding graphical representation is done for a camera-based offline video that contains handwritten mathematical expressions.

Key words: Key-frame, binarization, seam-line carving, gmentation, gabor-filter

INTRODUCTION

Mathematics is used in areas like education, business, management, research and almost all. So extraction of the mathematical expression is very essential. This study focuses on extracting the mathematical expression from a camera-based offline video. The challenges that aroused were finding the key-frames as key frame gives complete information of the video, proper binarization and calculating the orientation of the formula because of handwritten expression. All these challenges have been overcome with suitable algorithms which are explained in the proposed work.

The need of character recognition is increasing day-by- day. It is not limited with recognizing languages like English, Devengiri, French but has stretched its hand in recognizing digits and mathematical expression. Hence, feature extraction and finding the best classifier is the crucial part in the overall work. This study has two parts: extraction and recognition of the mathematical expression and building an automatic xy-plane to represent the recognized expression. This was developed for the easy understanding of the physical meaning of every equation, i.e. to give a visualization of the recognized expression in the physical world.

Literature review: The related works are done in recognizing mathematical expression both offline and online. In both the modes, the work is being increasing at a higher rate. The different stages in extracting the mathematical formulae are key-frame extraction, binarization, text line segmentation, character segmentation, feature extraction, classification and finally representing the recognized expression.

Key-frame is a frame that has the important information about the video. The characteristics of a video are described by the extracted Key-frames. Key-frames give the complete details of a video. In case of key-frame extraction, Guozhu Liu explains extraction of the key-frame from a video stream of MPEG (Liu and Zhao, 2010). Shot activity based key-frame extraction is explained by Kamoji *et al*. (2014). Here the authors have computed the reference histograms and activity indicator. The local minima are selected on the activity curve which constitutes the key frames. Liu *et al*. (2003) have proposed the Motion analysis based key frame extraction. Here, optical flow was computed for each frame and then a simple motion metric was used to identify the changes in the optical flow. The place where the local minima are found by identifying the change is marked as key frames.

In case of binarization, mostly used binarization method is Thresholding, i.e. given a grayscale or a RGB image a threshold is fixed. Pixels greater than the threshold is considered as foreground and pixels less than the threshold are considered as background. One way of fixing the threshold is taking the mean gray scale value of the image. This works better for scanned documents. But for camera based images several algorithm has been proposed. Stathis *et al*. (2008) have proposed an evaluation technique for binarization algorithms where

Corresponding Author: R. Suganya, Department of Computer Science and Engineering, Amrita School of Engineering, Coimbatore, India

different binarization methods suchas Otsu Thresholding, k-means, histogram peak, Pun-Thresholding and local Thresholding algorithms.

In text line segmentation, horizontal histogram profiling is used mostly for printed documents. For hand written characters profiling do not give satisfying results. So, many techniques have been implemented for text line segmentation for hand written characters. There are three ways in which text line segmentation is done. Top-down Bottom-up and Hybrid (Awal *et al*., 2009; Saabni and ElSana, 2011). The candidate text lines location is estimated first in Top-Down approach. The higher probability text lines are assigned to components thereby refining the estimation. Multiple text lines that touch each other are eliminated by splitting the large components. On the other hand in bottom-up approach local components are found first which are the connected components always and based on different grouping algorithms the components are grouped together to form a separate text line. Example Pixels to components, Components to words. Hybrid is the combination of both. Feature extraction and classification are the crucial parts in all recognition systems.

For handwritten characters the features extracted are mostly zoning features, Gabor-filter filter and structural features. Arya *et al*. (2015) proposed their research in recognition of Devnagari characters using Gabor-Filters. Murthy and Hanmandlu (2011) tell feature extraction using zoning features.

Classification is the method of putting something that belongs to the same category depending upon the features that are extracted. Two types of classifiers) Supervised Classifiers and Unsupervised Classifiers. Supervised classifiers classify depending on the training samples provides. On the other hand unsupervised classifiers adopt themselves to the situation and classify depending on the weight or measures given. For character recognition both the classifiers are used. Mostly used classifiers are K-nn classifier (Murthy and Hanmandlu, 2011), ANN classifier and Neural networks with back propagation method. Template matching is also done for character recognition. Visualization of the expression is the graph in xy-plane as how it can be expressed and viewed in the physical world. It is usually done by drawing and making models corresponding to the expression.

MATERIALS AND METHODS

System architecture: The overall view of the proposed system is described in Fig. 1. Each of the modules is explained.

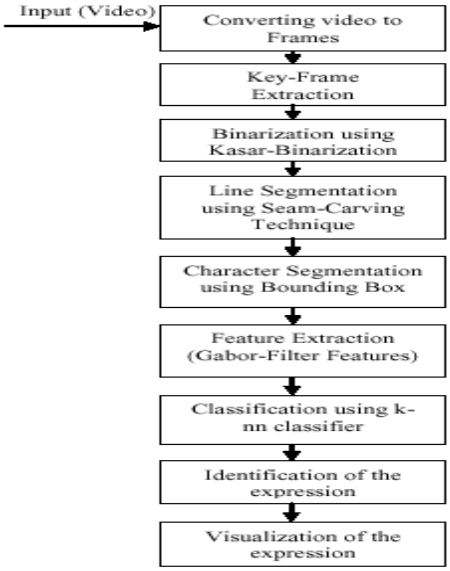

Fig. 1: System overview

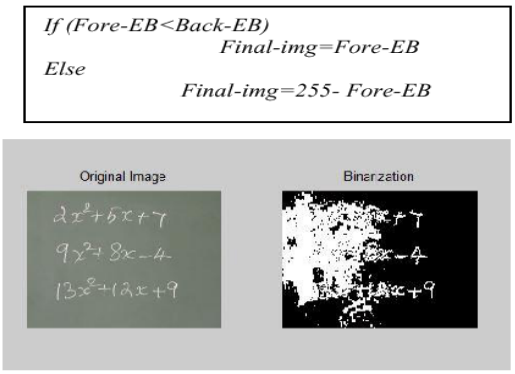

Fig. 2: Effect of thresholding

Identifying the key frames: Before identifying the key frames, the video is converted into frames. The key frames are identifies by taking the histogram difference between every nth and (n-1) th frame. The n th frame is identified as the key frame if the difference is greater than the threshold.

Binarization: Since, a camera based image is identified as a key-frame, the effect of threshold method is shown in Fig. 2. So, the binarization technique used is Kasar-binarization (Kasar *et al*., 2007; Kasar and Ramakrishnan, 2007). Figure 3 shows the output obtained using Kasar-binarization algorithm. This method can be directly applied on a RGB image instead of gray-scale. Steps involved in this technique are as follows:

- Find Canny-edge detection in R-plane (ER), G-plane (EG) and B-plane (EB) separately

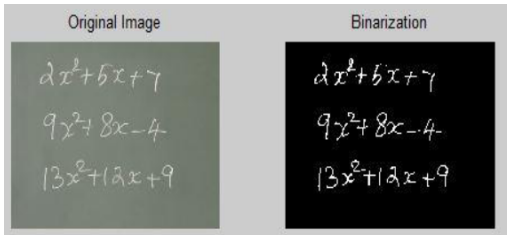

Fig. 3: Effect of Kaser-Binarization

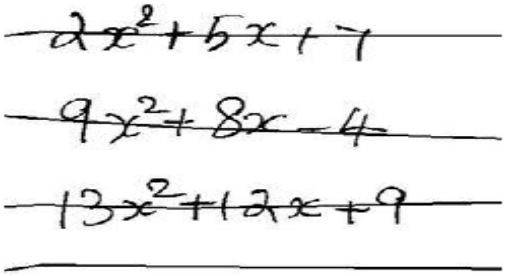

Fig. 4: Medial Seam

- Using OR-operator, (i.e.) E = ER V EG V EB, the edge image E is derived. Now E is used for further processing
- In the edge image E, connected component analysis and bounding box is done. Thus the obtained is the Edge Box (EB). The aspect ratio of Edge Box is maintained between 0.1 and 10 to filter out the elongated and non text regions. The EB having pixels >15 pixels and less than the 1/5th of the size of image is taken for further processing
- The Foreground intensity (Fore-EB) is calculated as the mean grayscale intensity of pixels corresponding to the Edge pixels. The background intensity (Back- EB) is calculated as median of the corner of the edge pixel. The final image is obtained as follows

Text line segmentation: The method used for Segmentation is Seam-Carving Technique. Seam is a junction or line where two pieces join, in other words, seam is a path of low energy pixels that are connected optimally. Seam carving method was initially used for Content-aware Image resizing, then video retargeting and Segmentation. There are two types of seams, vertical and horizontal seam (Zhang and Tan, 2014; Saabni and ElSana, 2011). For our method we go for horizontal seam. The approach is in finding the energy of the pixel. The energy of the pixel decides whether the pixel is to be included or not. High energy is mainly the text area and low energy is the background. Here the work goes in mainly calculating.

- Medial seam (Fig. 4) is a path that crosses along the text area

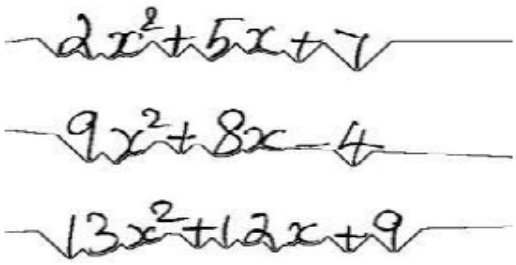

Fig. 5: Separating seam

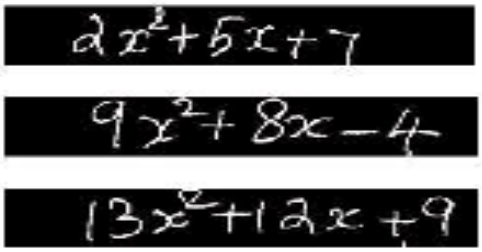

Fig. 6: Segmented text lines

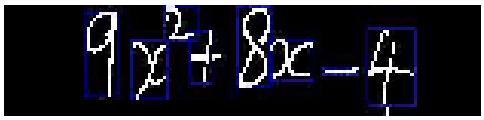

Fig. 7: Character segmntation

- Separating seam (Fig. 5) is a path that passes between two consecutive text lines (Saabni and ElSana, 2011)

The separating seam is found by the energy map. The Separating seam has (no-of-lines X no-of-columns) values corresponding to the coordinates of the image. To segment each expression line by line, the starting pixel of the i th line is minimum value of (i-1) th seam line and the ending pixel is the maximum of the ith seam line. By identifying this If (Fore-EB<Back-EB) Final-img = Fore-EB Else Final-img = 255- Fore-EB each line is segmented. Figure 6 represents the segmented text lines.

Character segmentation: Each character is segmented using 8-component bounding box analysis. Each character is now segmented and stored. Figure 7 represents the bounding box segmentation of characters.

Feature extraction: Features are the ones that give complete information of the image. For feature extraction Gabor-Filter features are used. From the Gabor-filter local energy and mean-amplitude is used as features, since it gives better results for handwritten characters. Figure 8 represents Gabor filter with different scaling and orientation.

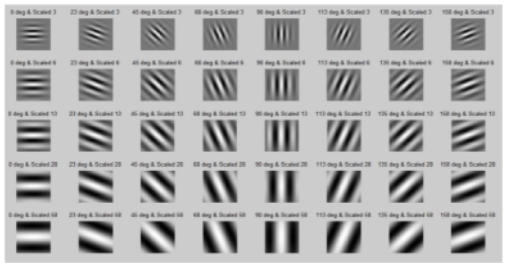

Fig. 8: Gabor filter with different scaling and orientation

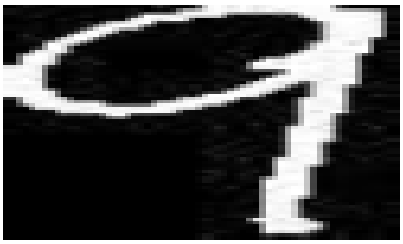

Fig. 9: Sample character

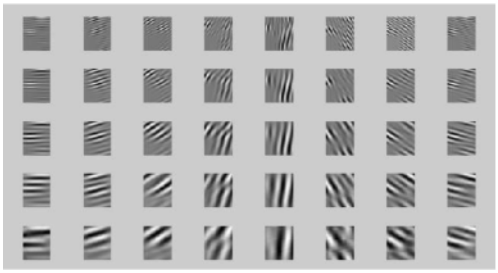

Fig. 10: Sample character convolved with gabor filter

By convolving the character with the derived 40 filters, [1×40] features are extracted for a single character. Figure 9 is the sample character and Fig. 10 shows the character convolved with the Gabor filters.

- Number of characters trained = 793
- Feature extracted for each Character = 40
- Totally, 31720(793×40) features have been extracted and these features have been used for training

Classification: Classification is the crucial part in all the recognition area. Here, for classification, a supervised k-nn classifier is used. K-nn classifier classifies based on the distance measures and the maximum votes it receives amongst its neighbors.

During classification, for the test image 40 Gabor filter features are extracted. The features extracted from the test image are compared with the features that have been trained. The distance is found between the test feature and the training features. The minimum distance is used for classification and the resultant is the training image that has the minimum distance with the test image. Once the characters are recognized, the expressions are stored in a text file. Figure 11 shows the formula that have been extracted and stored as a text file.

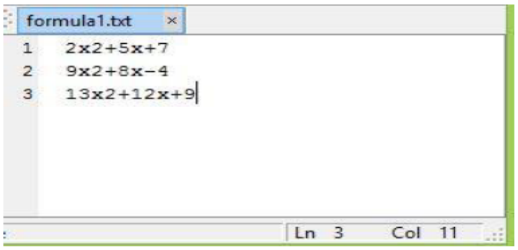

Fig. 11: Formula extracted and stored as a text file

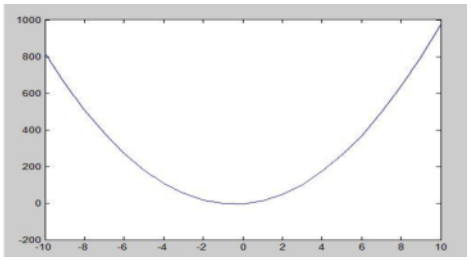

Fig. 12: Representation the recognized expression 9×2+8x-4

Table 1: Count of classes

Character	Count
Alphabets [a-z]	26
Digits [0-9]	10
Special symbols [+,-,*,/,(,)]	06
Total	42

Table 2: Count of characters for each class

Character	Count
Alphabets [a-z]	413
Digits [0-9]	260
Special symbols [+,-,*,/,(,)]	120
Total	793

Representation of the expression: Since in this study, it is concerned only about the polynomial expressions, an automatic xy-plane is generated and the recognized expression is represented in the generated xy-plane that gives the graphical representation of the recognized expression. Figure 12 gives the representation of the equation 9×2+8x-4.

Data set details: The video taken is at a rate of 60 fps (frames per second). The video consists of handwritten mathematical expression. For training 793 characters have been written and trained. Characters includes [a-z, 0-9, +, -*, /, (,)]. Software used is MATLAB-2013a. Table 1 shows the count of classes. Table 2 gives the details regarding the characters for each class that have been used for training.

Table 3: Confusion matrix

Predicted			
N = 96	Positive	Negative	Total
Actual			
Negative	6 (a)	3 (b)	9
Positive	2 (c)	85 (d)	87
Total	8	88	96

RESULTS AND DISCUSSION

Satisfactory results have been achieved through this work. The binarization approach, text line segmentation, character segmentation has been the base for the character recognition process.

- Total number of class = 42
- Number of expression for testing = 10
- Total number characters present in each symbol (N) = 96

The following table Table 3 shows the confusion matrix for character recognition process. The following terms has been calculated based on the entries of the confusion matrix:

- Accuracy = (a+d)/(a+b+c+d) = 0.947
- Recall = d/ (c+d) = 0.977
- Precision = d/(b+d) = 0.965

This results has been obtained with k-nn classifier having value of k = 2 and distance as Euclidean distance.

CONCLUSION

Recognition of handwritten mathematical expression from a camera-based video and to build the xy-graphical representation for the recognized expression is carried out in this study.

RECOMMENDATIONS

This study concentrates only on two-dimensional equations. Future work can be carried with all types of mathematical equations and also for real time videos. Figure 11 formula extracted and stored as a text file.

REFERENCES

Arya, S., I. Chhabra and G.S. Lehal, 2015. Recognition of devnagari numerals using gabor filter. Indian J. Sci. Technol., Vol. 8, 10.17485/ijst/2015/v8i27/81856.

Awal, A.M., H. Mouchere and V.C. Gaudin, 2009. Towards handwritten mathematical expression recognition. Proceedings of the 2009 10th International Conference on Document Analysis and Recognition, July 26-29, 2009, IEEE, New York, USA., ISBN:978-1-4244-4500-4, pp: 1046-1050.

Kamoji, S., R. Mankame, A. Masekar and A. Naik, 2014. Key frame extraction for video summarization using motion activity descriptors. Int. J. Res. Eng. Technol., 3: 491-495.

Kasar, T. and A.G. Ramakrishnan, 2007. Specialized text binarization technique for camera based document images. WISP., 2007: 28-31.

Kasar, T., J. Kumar and A.G. Ramakrishnan, 2007. Font and background color independent text binarization. Proceedings of the Second International Workshop on Camera-Based Document Analysis and Recognition, September 22, 2007, Indian Institute of Science, Bangalore, India, pp: 3-9.

Liu, G. and J. Zhao, 2010. Key frame extraction from MPEG video stream. Proceedings of the Third International Symposium on Information Processing (ISIP.), October 15-17, 2010, IEEE, New York, USA., ISBN:978-1-4244-8627-4, pp: 423-427.

Liu, T., H.J. Zhang and F. Qi, 2003. A novel video key-frame-extraction algorithm based on perceived motion energy model. IEEE Trans. Cir. Syst. Technol., 13: 1006-1014.

Murthy, O.R. and M. Hanmandlu, 2011. Zoning based Devanagari character recognition. Int. J. Comput. Appl., 27: 21-25.

Saabni, R. and J. ElSana, 2011. Language-independent text lines extraction using seam carving. Proceedings of the 2011 International Conference on Document Analysis and Recognition, September 18-21, 2011, IEEE, New York, USA., ISBN:978-1-4577-1350-7, pp: 563-568.

Stathis, P., E. Kavallieratou and N. Papamarkos, 2008. An evaluation technique for binarization algorithms. J. Universal Comput. Sci., 14: 3011-3030.

Tapia, E. and R. Rojas, 2003. Recognition of on-line handwritten mathematical expressions using a minimum spanning tree construction and symbol dominance. Proceedings of the International Workshop on Graphics Recognition, July 30-31, 2003, Springer, Berlin, Germany, PISBN: 978-3-540-22478-5,-pp: 329.

Zhang, X. and C.L. Tan, 2014. Text line segmentation for handwritten documents using constrained seam carving. Proceedings of the 2014 14th International Conference on Frontiers in Handwriting Recognition (ICFHR.), September 1-4, 2014, IEEE, New York, USA., ISBN:978-1-4799-4335-7, pp: 98-103.