Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No. 5 (2023)

# Performance Analysis of Color K-Means andRange Filter for Text Detection in Images or Video

**[1]Dayananda K J, [2]Dr. Puttegowda D**

[1]Department of Computer Science and EngineeringATME College of Engineering, Mysuru, India
Visvesveraya Technological University,
Belgavi-590018
[2]Department of Computer Science and EngineeringATME College of Engineering, Mysuru, India
Visvesveraya Technological University,
Belgavi-590018

**Abstract—**Text extraction in images and videos is an imperative role in image study and computer vision. The text identification process determines the presence of textual contents in the given input. The main challenges that occurred during the text identification process are complex background, illumination, low-resolution, different color, and variation in font size. In this paper, the color k-means is employed to raise the gap among textual and non-textual information. A range filter is employed to identify possible text components. Aperture concept has established to identify the actual text sections. The accuracy of the provided model was estimated using the common datasets MRRC, hua's, and nus dataset.. The outcome shows that the approach is promising and encouraging.

**Keywords-** Color K-means; Range filter; Aperture; text detection;

## 1. Introduction

A text is anything that conveys a meaning to the person and can be read.. Videos' text often contains a wealth of insightful information. Text extraction from images and videos is a challenging research field that fascinates a huge number of scholars. Although various character recognition techniques have been developed, the problem of text detection has not been thoroughly addressed. Text detection determines the existence of texts where text detection detects the region of text pixels in an image and videos. The main challenges arrived in the text detection are complex background, low resolution, alignments of characters, lighting and varied shapes, size and color of fonts. The motivation of text detection is an attempt towards the development of arbitrary oriented multilingual text detection. The Previous models were able to detect the text in a clear background with low resolution, but most of the models failed to detect the text's present illumination situation.. This research work is able to detect the text in illumination situationwith considering all kinds of challenges.

As we have seen in this work, the text region has distinguishing characteristics from non-text regions, and connectivity is separate from its background. The input image was divided into a set of colors using the color k-means method in the first stage. Range filter makes edges of the image perceptible.The aperture concept is employed on the filtered image to extract the true text contents. By looking at the bounding box region of text components and non-text components, false positives are then removed. By using k-means, range filter, and aperture feature we presented a text identification model for images and videos. Investigational results proved that the presented model efficiently detect the region of text pixels in the image. The proposed model is robust for font-color, font-size, background complexity and illumination.

## 2. Related Works

An analytical company that develops a trial to develop a computer architecture with the ability to automatically scan from images and videos the text information visually presented in challenging backgrounds is known as text detection in images and videos. Due to the variation in sizes, colors, and backdrop complexity, automatic character detection in video sequences can be a challenging problem.

Shin et.al proposed support vector machine-based text detection in cardinal video by using controlled texture arrangement and SVM for texture classification. Window concept is used to distinguish the text pixel

from a background [1]. Chen et.al developed text identification model using SVM and MLP concept [2]. Chen et al introduced text detection and verification model. For the purpose of finding the candidate text regions, a quick method is used. Contrast features were employed to train the machine during the verification stage. [3]. Jiang et al presented a learning-based model to isolate the text information in the scene images. Niblack clustering algorithm extracts theconnected-components. Two-stage classification module is applied to verify the text candidates [4]. Ye and Huang developed a new model for text detection using texture features and SVM. This model is high detection rate and low false alarm rate [5]. Ye et al presented a morphological operation, wavelet feature and SVM concept to identify the text region [6]. Based on a spatial restoration and temporal restoration model, Lee et al. created an automatic text detection method. To distinguish the text pixels from the background, the SVM model is used [7]. A text information extraction approach based on the mean horizontal run length notion was presented by Nor et al [8]. Liu etal applied edge detection technique, texture feature, k-means and empirical rules to identify the true text information [9]. Ma et al proposed a technique to detect arbitrary-orientated text information. This technique developed using connected components and top-down pruning procedure [10]. According to the literature review, there is a need for an effective and efficient text detection approach. In order to identify the text region, this paper has provided the range filter using aperture concept.

## 3. Proposed Methodology

To separate the text from images and videos, a new text identification technique was invented[1]. The proposed conceptual consists of color k-means clustering, range filter and finding aperture and then considering it asa text. K-means tries to cluster data based on similarity. The range filter identifies the observable image's boundaries and outlines. Aperture is a technique in which a text is represented by a hole in the image. In the aperture notion, the beginning and finishing positions are identical[12-13]. Finally, background false positives that are present are eliminated. A flow diagram of the suggested methodology is shown in Fig. 1.

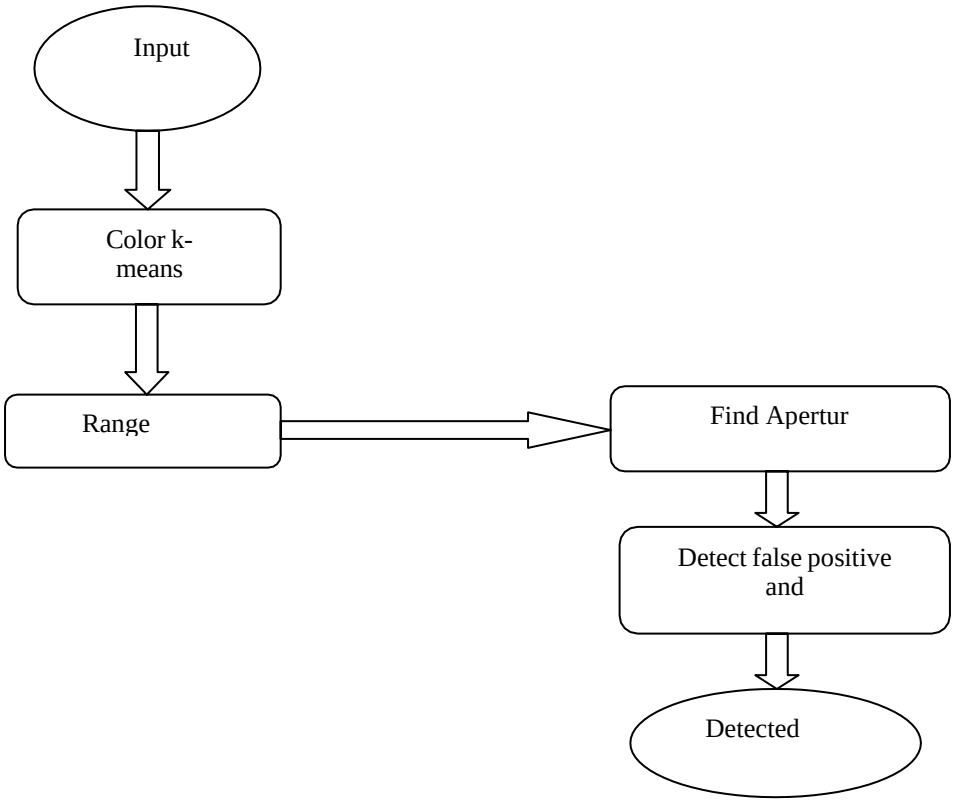


**Fig.1.** Flow Chart Of The Proposed Methodology

### A. Color K-Means

The color K-means approach has employed to group the pixel into a related region for RGB image to get the effectiveness compared to the gray image. The color k-means cluster requires parameter k as part of its input. It exactly divides the image into a given set of colors.

- It computes the mean of each one of the clusters and calculates the space of each point from theequivalent cluster mean.
- Iterated the steps until a summation of squared within collection of errors not lowered anymore.

Iteration tries to decrease the summation, overall collections of the squared group faults where the distance of a point to the respective group means. It takes the distance from the point of the data to prototype as the objective function of optimization. Color k-means produce higher clusters and compute fast when variables are huge. Fig.2shows the color k-means output.

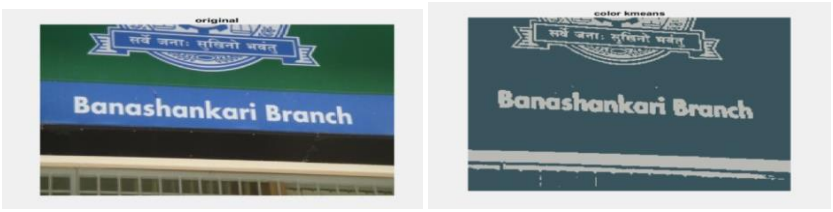


**Fig.2.** Input And Result Of Color K-Means

### B. Range Filter

The range filter identifies the edges and outlines of the image. This technique works based on the texture features. Fig.3 shows the output of the range filter. It can also use to find texture boundaries. Range filter technique is applied a 3X3 neighborhood concept. The range filter is the difference between Max filter and Min filter. A rangefilter is used to find edge and borders in the binary image. Range filter will find a pixel intensity value within windows is described. The edge detection properties of range filter can be illuminated using noiseless images of binary images which gives the proper output of edge. By selecting the parameter of the range filter the output may be positioned on the light or the dark side of the edge.


**Fig.3.** Result Of Range Filter.

### C. Aperture

The aperture concept uses an A range as a text and the identical starting and finishing points. Imfill is used to locate the hole in the input image's connected component, and the absolute difference between the connected component and the hole component is then calculated. More number of holes are generated for text components as compared to the background. Later, By bridging the interior gaps of the holes, imfill aids in identifying the whole region of text components. After finding the aperture if false positives occur. False-positive can be eliminated by finding zero and non-zero elements. If the number of zero elements is zero then boundaries are drawn for the text. The aperture in images and videos is used to identify text candidates. Fig.4 final result of the suggested technique.

## 4. Experimental Results

Experimentation conducted on a standard dataset of natural images. Here, To evaluate the effectiveness of the suggested model, three datasets were taken into consideration.. These three datasets comprises all kinds of challenges with more number of samples. Hua's datasets include 45 images that have only horizontal text with complex background and achieved precision 63.2%, recall 72.5% and f-measure 67.5%. MRRCT includes standard 167 images which are multilingual text and achieved precision 68.1%, recall 71.3% and f-measure 69.6%. NUS dataset includes 113 images that have caption text and achieved precision 70.8%, recall 72.4% and f-measure 71.5%. Fig.5 shows the MRRCT results. Fig.6 shows Hua's dataset results. Fig.7 shows Nus dataset results. Theequation 1, 2 and 3 shows the precision, recall and f-measure formulas. Here actual text box (ATB) means the entire number of text lines present in the input, total text box (TTB) means text boxes detected by the proposed method and false box (FB) means the non-text region is detected as a text box.

Precision=Actual Text Box/(Total Text Box + False Box) (1) Recall=Actual Text Box/ Total Text Box (2)
F-measure=(2*Precision*Recall)/(Precision + Recall) (3)


**Fig.5.** Results Of The Mrrct Dataset.


**Fig.6.** Shows The Results Of Hua's Dataset.


**Fig.7.** The Nus Dataset's Result.

## 5. Conclusion

The existence of text in images or video provides rich semantic information. The text detection process finds the text region using range filter and aperture concept. In this work, initially, The link between text pixels

and non-text pixels is improved using the color k-means method. Later, the range filter extracts the possible text components from the resultant of the color k-means algorithm. The true text candidates in the input can be found using the aperture principle. False alarms are eliminated through the use of heuristic rules. Finally, By running the experiment on three different datasets, the effectiveness of the approach is assessed.. In future work, an efficient cluster method and text pixel distribution model will be developed to differentiate the text component from the non-text component.

**References**

[1] Shin, C. S., Kim, K. I., Park, M. H., & Kim, H. J. (2000, December). Support vector machine-based text detection in digital video. In Neural Networks for Signal Processing X. Proceedings of the 2000 IEEE Signal Processing Society Workshop (Cat. No. 00TH8501) (Vol. 2, pp. 634-641). IEEE.

[2] Chen, D., Odobez, J. M., & Bourlard, H. (2004). Text detection and recognition in images and video frames. Pattern recognition, 37(3), 595-608.

[3] Chen, D., Odobez, J. M., & Thiran, J. P. (2004). A detection/verification scheme for finding text in images and video frames based on contrast independent features and machine learning methods. Signal Processing: Image Communication, 19(3), 205-217.

[4] Jiang, R. J., Qi, F. H., Xu, L., Wu, G. R., & Zhu, K. H. (2007). A learning-based method to detect and segment text from scene images. Journal of Zhejiang University-Science A, 8(4), 568-574.

[5] Ye, Q., & Huang, Q. (2004, November). A new text detection algorithm in images/video frames. In Pacific-Rim Conference on Multimedia (pp. 858-865). Springer, Berlin, Heidelberg.

[6] Ye, Q., Gao, W., Wang, W., & Zeng, W. (2003, December). A robust text detection algorithm in images and video frames. In Fourth International Conference on Information, Communications and Signal Processing, 2003 and the Fourth Pacific Rim Conference on Multimedia. Proceedings of the 2003 Joint (Vol. 2, pp. 802-806). IEEE.

[7] Lee, C. W., Jung, K., & Kim, H. J. (2003). Automatic text detection and removal in video sequences. Pattern Recognition Letters, 24(15), 2607-2623.

[8] Nor, D. M., Omar, R., Jenu, M. Z. M., & Ogier, J. M. (2011). Image Segmentation and text Extraction: Application to the Extraction of Textual Information in Scene Images. In International Seminar on Application of Science Mathematics (ISASM2011).

[9] Liu, C., Wang, C., & Dai, R. (2005, August). Text detection in images based on unsupervised classification of edge-based features. In Eighth International Conference on Document Analysis and Recognition (ICDAR'05) (pp. 610-614). IEEE.

[10] Ma, L., Wang, C., & Xiao, B. (2010, October). Text detection in natural images based on multi-scale edge detection and classification. In 2010 3rd International Congress on Image and Signal Processing (Vol. 4, pp. 1961-1965). IEEE.

[11] Puttegowda, D., & Padma, M. C. (2016, August). Human Motion Detection and Recognising their Actions from the Video Streams. In Proceedings of the International Conference on Informatics and Analytics (pp. 12).

[12] Basavaraju, H. T., Aradhya, V. M., & Guru, D. S. (2018). A novel arbitrary-oriented multilingual text detection in images/video. In Information and decision sciences (pp. 519-529). Springer, Singapore.

[13] Dayananda, K. J., & Puttegowda, D. (2019). A Comprehensive Study on Text Detection in Images and Videos. Journal of Innovation in Computer Science and Engineering, 9(1), 17-23.