

Color K-means, Gaussian Filter and Aperture Concept for Text Localization in Images

K. J. Dayananda¹ and D. Puttegowda²

¹⁻²Department of Computer Science and Engineering ATME College of Engineering, Mysuru, India
Email: dayananda.kem@gmail.com, puttegowda.77@gmail.com

Abstract—Text extraction in image is an essential role in machine learning and computer vision field. The text localization process determines the presence and location of text in the given inputs. The challenges, which are occurred during the text localization task is different orientation, low-resolution, complex background and illumination with variation in font size and color. In this research paper, the color k-means is applied to make separate group of colors present in the text image. Histogram equalization process is applied to sharpen the text pixels from the background pixels. The Gaussian filter is applied to combine all text pixels together. Aperture concept has implemented to locate the actual text regions. The standard datasets like hua's and nus dataset were used to estimate the performance of the presented model. Precision, recall and f-measure is used to estimate the performance of the proposed model.

Index Terms— Color K-means, Histogram Equalization, Gaussian filter, Aperture, Text localization.

I. INTRODUCTION

A text information conveys a set of meanings to the person which can be read and understand easily. Text in videos carries rich set of information. Text extraction from images are challenging research field that fascinates a huge number of scholars. Although various optical character recognition is developed still the problem of text localization is not thoroughly solved. Text localization determines the existence of texts in a particular place, where text localization identify the region of text pixels in an image. The main challenges arrived in the text localization are complex background, low resolution, alignments of characters, lighting and varied shapes, size and color of fonts. The motivation of text localization is an attempt towards the development of arbitrary oriented multilingual text localization. The previous models are able locate the text in clear background with low resolution, but most of the models fails to locate the text present illumination situation. This research work is able to locate the text in illumination situation with considering all kinds of challenges.

In this work, as we have observed the text region has distinctive features compared to non-text regions and connectivity is different from its background. In the first step, we used a color k-means algorithm divides the input image into a set of colors given by k as a parameter. Then histogram equalization and Gaussian filter is used to perceive the regions of texture in an image with the help of the texture filter function. Gaussian filter makes edges of the image perceptible. The aperture concept is employed on the filtered image to extract the true text contents. Then false positives are eliminated by studying the bounding box region of text components and non-text components. By using k-means, Gaussian filter, and aperture feature we presented a text identification model for images and videos. Investigational results proved that the presented model efficiently locate the region

of text pixels in the image. The proposed model is robust for font-color, font-size, background complexity and illumination.

II. LITERATURE SURVEY

Text localization in images and videos can be an analysis house that produces a trial to grow a computer structure with the facility to robotically scan from pictures and videos the text information visually presented in difficult backgrounds. Automatic character localization in video sequences can be an advanced task, thanks to the variability of sizes and colors additionally on the complexity of the background.

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 the background [1]. Chen et.al developed text identification model using SVM and MLP concept [2]. Chen et al introduced text detection and verification model. A fast algorithm is employed to identify the candidate text regions. In the verification stage, contrast features were used to train the machine [3]. Jiang et al presented a learning-based model to isolate the text information in the scene images. Niblack clustering algorithm extracts the connected 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]. Lee et al developed an automatic text detection approach based on spatial restoration and temporal restoration model. SVM model is applied to separate the text pixels from the background [7]. Nor et al introduced a text information extraction model using mean horizontal run length concept [8]. Liu et al 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]. Huang et al [11] proposed new approach for text localization in natural images, by discriminating text and non-text regions at three levels: pixel, component and text line levels. Mengdi et al [12] extract text information effectively from natural scene image with complex background, multi-orientation perspective and multilingual languages, we present a new method based on the improved Stroke Feature Transform (SWT). Alowaidi and Alshehri [13] developed a domain adaptation network (DAN) to learn the domain shift from day images to night images. This proposed sub-network DAN can be attached to any existing state-of-the art text detector to learn the domain shift from day to night. Chaitra and Dinesh [14] proposed a detailed exploration of various filter techniques such as radon transform, conservative filter, Gabor filter, bilateral filter, and context-aware filters with maximally stable extremal region for text localization. As per the literature review, there is demand for effective and efficient method for text detection. Hence, this paper has presented histogram equalization and Gaussian filter with aperture concept to identify the text area.

III. PROPOSED METHODOLOGY

A new text identification method has developed to isolate the text information in images and videos [15]. The proposed method consists of color k-means clustering, histogram equalization, Gaussian filter and finding aperture and then considering it as a text. K-means tries to cluster data based on similarity. The Gaussian filter identifies the edges and outlines of the image observable. Aperture is which consists of a hole in the image and that hole is considered as a text. The aperture concept is where the starting point and ending points are the same [16]. Finally, false positives are removed which is present in the background. Fig.1 shows a 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 the equivalent 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

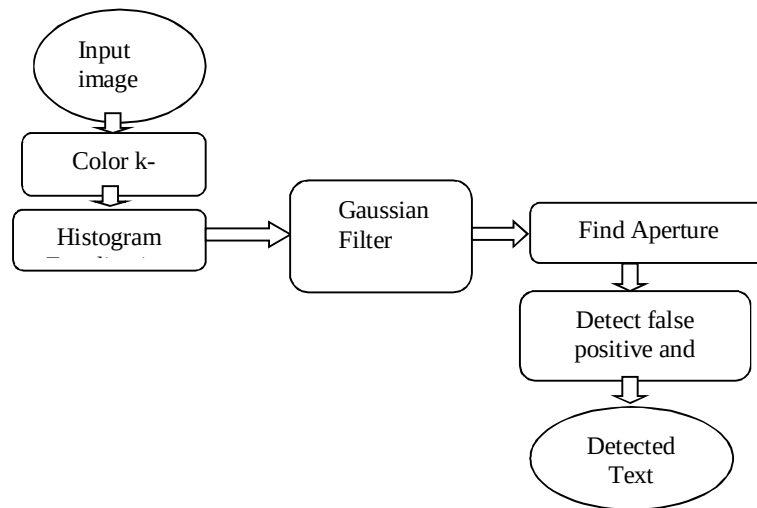


Figure 1. Flow chart of the proposed methodology

Color k-means produce higher clusters and compute fast when variables are huge. Fig.2 shows the color k-means output.

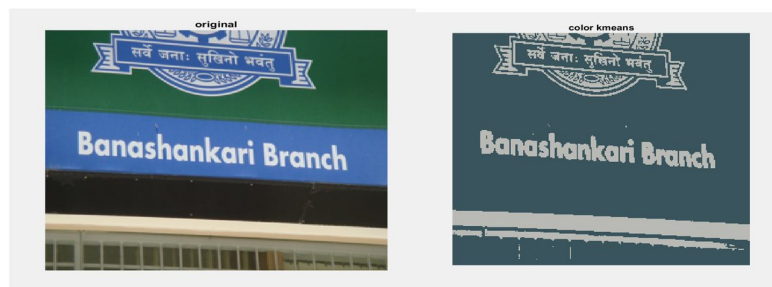


Figure 2. Input and Result of color k-means.

B) Histogram Equalization and Gaussian Filter

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



Figure 3. Result of histogram equalization and Gaussian filter.

C) Aperture

The aperture concept is where the starting point and ending points are the same and which A range as a text. Imfill is used to find the hole from the connected component of the input image, then the absolute difference is taken from the connected component and the hole component. More number of holes are generated for text components as compared to the background. Later, imfill helps to find out the entire region of text components by filling the inside gap of the holes. 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. Text candidates are identified by the aperture in the images and videos. Fig.4 final output of the proposed method.



Figure 4. Final Result of aperture after removing false positives

V. EXPERIMENTAL RESULTS

Experimentation conducted on a standard dataset of natural images. Here, three datasets have been considered to measure the performance of the proposed model. These three datasets comprise all kinds of challenges with a greater number of samples. Hua's datasets include 45 images that have only horizontal text with complex background and achieved precision 67.4%, recall 75.3% and f-measure 71.1%. NUS dataset includes 113 images that have caption text and achieved precision 71.3%, recall 77.2% and f-measure 74.1%. Fig.5 shows Hua's dataset results. Fig. 6 shows Nus dataset results. The equation 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.

$$Precision = \frac{Actual\ Text\ Box}{(Total\ Text\ Box + False\ Box)} \quad (1)$$

$$Recall = \frac{Actual\ Text\ Box}{Total\ Text\ Box} \quad (2)$$

$$F - measure = \frac{2 * Precision * Recall}{Precision + Recall} \quad (3)$$



Figure 5. Shows the Results of Hua's dataset.



Figure 6. Results of the Nus dataset.

VI. CONCLUSION

The presence of text in images or video provides rich semantic information. The text localization process finds the text region using histogram equalization, Gaussian filter and aperture concept. In this work, initially, the color k-means algorithm helps to rise the relationship among text pixels and non-text pixels. Later, the histogram equalization and Gaussian filter extracts the possible text components from the resultant of the color k-means algorithm. The aperture concept has been discovered to detect the true text candidates in the given input. Heuristic rules are applied to remove false alarms. Finally, the performance of the method is evaluated by conducting the experimentation on three different datasets. 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., and 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., and 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., and 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., and 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., and 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., and 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., and 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., and 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., and 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] Huang, W., Lin, Z., Yang, J., and Wang, J. (2013). Text localization in natural images using stroke feature transform and text covariance descriptors. In *Proceedings of the IEEE international conference on computer vision* (pp. 1241-1248).
- [12] Jiang, M., Cheng, J., Chen, M., and Ku, X. (2018). An improved text localization method for natural scene images. In *Journal of physics: conference series* (Vol. 960, No. 1, p. 012027). IOP Publishing.