An Application of *K*-Means Clustering for Improving Video Text Detection

V.N. Manjunath Aradhya and M.S. Pavithra

Abstract. In the present work, we explore an extensive applications of Gabor filter and *K*-means clustering algorithm in detection of text in an unconstrained complex background and regular images. The system is a comprehensive of four stages: In the first stage, combination of wavelet transforms and Gabor filter is applied to extract sharpened edges and textural features of a given input image. In the second stage, the resultant Gabor output image is grouped into three clusters to classify the background, foreground and the true text pixels using *K*-means clustering algorithm. In the third stage of the system, morphological operations are performed to obtain connected components, then after a concept of linked list approach is in turn used to build a true text line sequence. In the final stage, wavelet entropy is imposed on an each connected component sequence, in order to determine the true text region of an input image. Experiments are conducted on 101 video images and on standard ICDAR 2003 database. The proposed method is evaluated by testing the 101 video images as well with the ICDAR 2003 database. Experimental results show that the proposed method is able to detect a text of different size, complex background and contrast. Withal, the system performance outreaches the existing method in terms of detection accuracy.

Keywords: Wavelet Transform, Gabor filter, *K*-means clustering, linked list approach, Wavelet Entropy.

V.N. Manjunath Aradhya
Department of Information Science and Engineering,
Dayananda Sagar College of Engineering, Bangalore
e-mail: aradhya.mysore@gmail.com

M.S. Pavithra
Department of Master of Computer Applications,
Dayananda Sagar College of Engineering, Bangalore
e-mail: pavithraspurthi@yahoo.com

A. Abraham and S.M. Thampi (Eds.): Intelligent Informatics, AISC 182, pp. 41–47.
springerlink.com © Springer-Verlag Berlin Heidelberg 2013

1 Introduction

Texture, Color and Shape based multimedia database registering and retrieving requires a task of text detection in images, video etc. Text detection is the process of determining the location of text in an image. Text region detection is mainly based on texture and dominant orientation. Text regions are detected either by analyzing the edges of the candidate regions or by using textural properties of an image. It provides primal information for text extraction and verification. Though many efforts have been devoted to, it remains a challenge due to variations of background, font of a text image. Text Information Extraction techniques can be broadly divided into two classes: i) region based and ii) texture based methods. In Region based methods connected components(CC) or edges are found on the basis of their perceptive difference with the background. This is followed by merging of the CCs or edges to get the text bounding boxes. In Texture based methods text in an image has distinct textural properties that gets distinguish from the background [1].

Most of the proposed text detection methods use text features, color, edge and texture information. So to extract a text from an image and discriminate it from the background, many researchers have applied heuristic rules based on empirical constraints and other few researchers have used machine-learning methods trained on real data. Aradhya et.al.[2] describe the text detection method using wavelet transform and Gabor filter. Kaushik et.al.[3] propose an approach for text detection, using morphological operators and Gabor wavelet. Phan et. al.[4] describe an efficient text detection based on the Laplacian operator. Shivakumara et.al.[5] proposed a Wavelet Transform Based technique for video text detection. Recently, few research works have carried out on K-means and connected component analysis in the domain of text detection in video images. Shivakumara et.al.[6] describe a method based on the Laplacian in the frequency domain for video text detection.

From the literature study, it is clear that, though the concept of K-means algorithm and a connected component analysis have used in many of the text detection approaches, the detection accuracy of the text region can still be improved without missing any data. By sustaining the development of the system [2], we propose a system with the combination of Gabor filter and K-means clustering is extensively used to detect the true text region accurately in attaining better detection rate with a very few missing data in numbers.

The remaining of our paper is structured as follows: In Section 2, as per the stages the proposed method is described. Section 3 presents the experimental results and performance evaluation on considered datasets, and finally in Section 4 conclusions are drawn.

2 Proposed Methodology

Proposed method is an improvised work of a robust multilingual text detection approach based on transforms and wavelet entropy [2]. An efficient texture feature

information is extracted by applying wavelet transforms and Gabor filter as described in [2]. The resulted Gabor output image is grouped into three clusters to classify the background, foreground and the sharpened texture edges obtained by applying the K-means clustering. In the next stage, morphological operations are performed to obtain connected components, then after a concept of linked list approach is in turn used to build a true text line sequence. In the final stage, wavelet entropy is imposed on an each connected component sequence in order to determine the true text region of an input image. The complete text detection procedure of our work is explained in the following subsections.

2.1 An Integrated Approach of Gabor Filter and K-Means Clustering for an Efficient Text Region Classification

The work performed using Wavelet transform and Gabor based method [2] is employed in our proposed method. In this we selected an average image of details of three orientation such as horizontal, vertical and diagonal images. The obtained detail information represents the sharpen edges of an image in all three orientations and this is subsequently used by Gabor filter to extract the textural information.

Gabor filter is optimally localized as per the uncertainty principle in both the spatial and frequency domain. That is the Gabor filter is highly selective in both position and frequency. This results in sharper texture boundary detection as in [2]. The main purpose of applying K-means clustering to a resultant Gabor image is to classify a highest energy class as a text candidates and the remaining classes as a uncertainty and non-text pixels. In the present work we considered three clusters to classify objects based on the feature set.

Choosing K is often an ad hoc decision based on prior knowledge, assumptions and practical experience [11]. Likewise we practically worked on choosing K value. Initially we set the value K=2 and observed true text pixels including false deteceted blocks. The obtained text classification result is shown in Figure 1(b) for the input image Figure 1(a). When we set the value K=3, we observed the true text regions are well classified compared to two clusters set in the first demonstration and results obtained to the same input image is shown in Figure 1(c) accordingly. The main idea behind applying K-means clustering by choosing a value K=3 for resultant Gabor image is, it considerably reduces the non-text pixels and efficiently classifies the text region from the background.

2.2 Morphological Operations and an Implementation of a Linked List Approach

To the obtained K-means clustering resultant image, we applied morphological operations to get a connected components of true text pixels. A concept of linked list approach [7] is then used to build a true text line sequence in order to get a sequence

Fig. 1 (a)Sample Input Image of 101 video images (b) K-means clustering images obtained for an input image when K=2 (c) K-means clustering image obtained for an input image when K=3

Fig. 2 (a) Resultant images obtained after applying morphological operations and as a sequence of true text line sequence of components after applying linked list approach (b) Resulted image of truly detected text region

of connected components to detect a sequence of true text regions of an input image. The results obtained for these stages are shown in Figure 2(a).

2.3 Wavelet Entropy

From the obtained sequence of connected components, we imposed the wavelet entropy to the corresponding region of a sequence of connected components in an input image, inorder to extract true text region as well to eliminate falsy blocks of an image. Then we extracted an energy information from an input image of the regions specified. Average energy of all the regions specified in the input image is fixed as threshold α. If the specified sequence of a text region $\geq \alpha$, where α is the threshold, it is considered as a text region or else considered as a non-text region. Figure 2(b) shows the text region obtained from the above mentioned procedure.

3 Experimental Results and Performance Evaluation

The proposed system is tested on two datasets. Firstly, a dataset of 101 video images provided by [4] comprising news programmers, sport video and movie clips. The dataset also includes both graphic text and seen text of different languages, e.g.

English, Chinese and Korean in the dataset. Second,the most cited ICDAR 2003 dataset [12], which contains images with text of varying sizes and positions. In order to evaluate the performance of proposed method, we used the following criteria:

- Truly Detected Block(TDB): A detected block that contains a text line, partially or fully.
- False Detected Block(FDB): A detected block that does not contain text.
- Text Block with Missing Data(MDB): A detected block that misses some characters of a text line (MDB is a subset of TDB).

Table 1 shows the results obtained for existing and proposed method on the database provided by [4]. For each image in the dataset we manually count the Actual Text Blocks(ATB). The performance measures defined as follows:

- Detection Rate (DR) = TDB / ADB
- False Positive Rate (FPR) = FDB /(TDB + FDB)
- Miss Detection Rate (MDR) = MDB / TDB

Table 1 shows the comparative study of proposed and existing methods. From this table is clear that the obtained TDBs are more i.e. the system detects more number of true text blocks, FDBs are sustained as in Transforms and Gabor based method, which indicates that there exists few alarms. MDBs are considerably reduced, which shows that the miss detection of text blocks are very few in number. We compared the proposed method with the existing text detection methods such as Edge-based [8], Gradient-based [9], Uniform-colored [10], Laplacian [4] and Transforms and Gabor based [2] methods. In order to evaluate the performance of the proposed method we considered 101 test images provided by [4]. From Table2, it is clear that the proposed method has higher DR and lesser MDR compare to existing methods and FPR is sustained as of Transforms & Gabor based method. The main goal of the proposed system is to achieve highest DR by detecting true text blocks of an image, we reached DR=98.9%, MDR=3.0% though FPR=13.7% which is sustained as of Transforms & Gabor based method. By the conduct of experiment, it is proved that the propose method exhibits higher detection rate and considerably lesser miss detection rate than the existing methods.

We also evaluated proposed method on a standard ICDAR 2003 dataset. Table3, shows obtained results and performance evaluation with the existing Transforms and Gabor based method on standard ICDAR 2003 dataset. The resultant text detection

Table 1 Results obtained for the dataset of 101 video imaages of[4]

Method	ATB	TDB	FDB	MDB
Edge-based[8]	491	393	86	79
Gradient-based[9]	491	349	48	35
Uniform-colored[10]	491	252	95	94
Laplacian[4]	491	458	39	55
Transform & Gabor based[2]	491	481	78	53
Proposed method	491	486	78	15

Table 2 Performance results obtained on dataset[8]

Method	DR	FPR	MDR
Edge-based[8]	80.0	18.0	20.1
Gradient-based[9]	71.1	12.1	10.0
Uniform-colored[10]	51.3	27.4	27.4
Laplacian[4]	93.3	7.9	7.9
Transform& Gabor based[2]	97.9	13.9	11.0
Proposed method	98.9	13.7	3.0

image of ICDAR 2003 dataset image is shown in Figure 3(b) for an input image shown in Figure 3(a). The vital part of our proposed method is that classifying the resultant Gabor image into three clusters by applying K-means clustering algorithm. With this we could able to detect true text regions effectively.

Table 3 Measures and Performance results obtained on ICDAR2003

Method	ATB	TDB	FDB	MDB	DR	FPR	MDR
Transform & Gabor based[2]	124	119	74	23	95.96	38.34	19.3
Proposed method	124	120	34	3	96.7	22	2.5

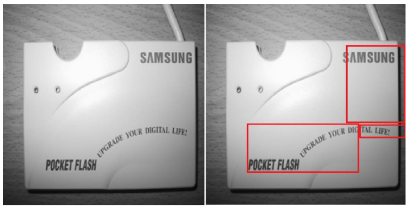

Fig. 3 (a) An input image of ICDAR 2003 dataset (b) The resulted text detection image

4 Conclusion

The proposed system is a development of an efficient text detection approach able to detect text of multilingual languages of different fonts, contrast and in unconstrained background. The key concept of our system is to detect true text region without missing any data, which is performed extensively by using the combination of Gabor filter and *K*-means algorithm. A concept of wavelet entropy which is used in our previous work [2] is applied to a result of the above mentioned combination of concepts to detect a true text region of an image. Experiments are conducted on two different datasets comprising of challenging images and varying background

images: (1) standard ICDAR 2003 dataset, (2) dataset of 101 video images. The present improvised proposed system performance analysis has done on dataset of 101 video images and standard ICDAR 2003 dataset. The proposed system exhibits better text detection with drastically decreasing the missing of data in a exact text region detection.

References

1. Chowdhury, S.P., Dhar, S., Das, A.K., Chanda, B., McMenemy, K.: Robust Extraction of Text from Camera Images. In: The Proceedings of 10th International Conference on Document Analysis and Recognition, pp. 1280–1284 (2009)
2. Manjunath Aradhya, V.N., Pavithra, M.S., Naveena, C.: A Robust Multilingual Text Detection Approach Based on Transforms and Wavelet Entropy. In: The Proceedings of Procedia Technology (Elsevier) 2nd International Conference on Computer, Communication, Control and Information Technology (C3IT-2012), Hooghly, West Bengal, India (in press, 2012)
3. Kaushik, N., Sarathi, D., Mittal, A.: Robust Text Detection in images using morphological operations and Gabor wavelet. In: The Proceedings of EAIT, pp. 153–156 (2006)
4. Phan, T., Shivakumara, P., Tan, C.: A Laplacian method for video text detection. In: The Proceedings of 10th International Conference on Document Analysis and Recognition, pp. 66–70 (2009)
5. Shivakumara, P., Phan, T., Tan, C.: A Robust Wavelet Transform Based Technique for Video Text Detection. In: The Proceedings of 10th International Conference on Document Analysis and Recognition, pp. 1285–1289 (2009)
6. Shivakumara, P., Phan, T., Tan, C.: A Laplacian Approach to Multi-Oriented Text Detection in Video. IEEE Transactions on Pattern Analysis and Machine Intelligence 33, 412–419 (2011)
7. Naveena, C., Manjunath Aradhya, V.N.: A linked List Approach for Handwritten Textline Segmentation. Journal of Intelligent Systems (in press, 2012)
8. Liu, C., Wang, C., Dai, R.: Text Detection in Images Based on Unsupervised Classification of Edge-based Features. In: ICDAR, pp. 610–614 (2005)
9. Wong, E.K., Chen, M.: A new robust algorithm for video text extraction. In: The Proceedings of First Asian Conference on Pattern Recognition, ACPR, vol. 36, pp. 1397–1406 (2003)
10. Mariano, V.Y., Kasturi, R.: Locating Uniform Colored Text in Video Frames. In: 15th ICPR, vol. 4, pp. 539–542 (2000)
11. Teknomo, K.: Kardi Teknomo's Tutorials (2006); Available via Kardi Teknomo, `http://people.revoledu.com/kardi/tutorial/kMean/index.html`
12. Lucas, S.M., Panaretos, A., Sosa, L., Tang, A., Wong, S., Young, R.: ICDAR 2003 Robust Reading competitions. In: ICDAR 2003, pp. 682–687 (2003)