ISSN (Print) : 0974-6846
Indian Journal of Science and Technology, Vol 8(15), 69986, July 2015 ISSN (Online) : 0974-5645

Hill Climbing Approach for Text Binarization from Videos

L. M. Merlin Livingston[1*] and M. Manikandan[2]

[1]Sathyabama University, Chennai - 600 119, Tamil Nadu, India; merlinlivingston@yahoo.com
[2]ECE Department, Madras Institute of Technology, Chennai, India; 2maniiz@yahoo.com

Abstract

Objectives: To develop presents an efficient algorithm for binarization of text from videos. However, it is a challenging and difficult task in image processing, due to their complicated background and non uniform character size. **Method/Analysis**: This study focuses a novel approach for binarization of text region from videos based on Hill climbing Algorithm. Symmetric filtering and graph cut algorithms are used to refine the obtained clusters. Finally, an optimal clustering selection algorithm is applied to obtain the text region. **Findings**: The experimental results show that the proposed text binarization technique is robust in text detection with various character size and complicated background. Data sets are taken from the You Tube Video Text (YVT) harvested from YouTube. **Application**: This proposed binarization technique is applicable to automated licence plate recognition system, especially to identify Indian vehicles licence plates.

Keywords: Binarization, Graph Cut Theory, Hill Climbing, K-means Clustering

1. Introduction

Text binarization from videos plays a vital role in computer vision problems. Even though the amount of video data available is rapidly increasing due to extensive use of camera phones, relatively little work has been done on extending text reading solutions to the video domain. Learning based document image binarization is described in[1]. Features such as mean and standard deviation for text region in document images are extracted using multilevel mapping and are used to train support vector regression. After feature learning, two binarization methods named grid-based Sauvola method and Lu's method are used for text binarization.

Kohonen Self-Organizing Map Neural Network (KSOMNN) based document binarization is presented in[2]. Initially, various binarization techniques such as Otsu, fuzzy C means, bernsen, Niblack are performed. Among them, best binarization algorithms are selected based on parameter set evaluation. Finally, the selected best parameter set values of each binarization algorithm are given to the KSOMNN for training and then independent binarization of document images is performed. Racing algorithm based document image binarization is implemented in[3]. Two statistical parameters based on objects by distance and Laplacian energy based technique combined with objects by distance is used for document image binarization. The evaluation is carried on old documents.

Diffusion based document image binarization is presented in[4]. Modified linear fusion approach is used to de-noise and then binarized the input image with extreme noisy condition. All the parameters are determined automatically usiZng a heuristic optimization algorithm. An approach for text detection, localization and extraction is described for multilingual including English and Chinese video[5]. Text detection is achieved by the sequential multiresolution paradigm, local thresholding of edge map, and the hysteresis text edge recovery. Text extraction is performed by adaptive thresholding, dam point labeling, and inward filling approaches. Binary texture analysis based binarization is reviewed for colour

* Author for correspondence

text images in[6]. It consists of four stages: dimensionality reduction, clustering, texture feature extraction, and selection of the optimal binary image. Binary texture features are extracted using run-length histogram and spatial-size distribution algorithms. Finally, optimal candidate is obtained by linear discriminate analysis classifier. An approach for image binarization is presented for degraded document images in[7].Initially, an adaptive contrast map is generated from input degraded document image. To identify the text stroke edge pixels, the constructed contrast map is binarized and combined with canny edge map. Finally, local threshold is estimated based on the intensities of detected text stroke edge pixels within a local window for document text segmentation. Connected operator based document image binarization is implemented in[8]. Component-tree paradigm approach is used that overcome the issue of binarization techniques; removal of small objects. An approach for text detection in indoor/outdoor images is discussed in[9]. Initially, Wiener filter is used to remove the noises and then rough estimation is done to identify foreground regions. Sauvola's approach is applied to gray scale and inverted gray scale image to obtain binary image. Metric based approach for colour text extraction is given in[10]. In order to locate text region, Log-Gabor filter is used to complement colour information to spatial and intensity information. A text validation measure is used to find text regions.

Adaptive graph cut based binarization is implemented in[11]. It is composed of three stages. At first, given text image region is adaptively splitting into several sub images, whereas canny edge detection and connected component analysis are employed. Then the polarity of the text region is determined thus text and background pixels with high accuracy are predicted. Subsequently foreground text region is segmented from the sub-images by using graph cut algorithm. Finally appropriate text region is obtained by merging segmented sub image. An approach using L1 -norm principal component analysis for text binarization is presented[12]. Both simple and complex background natural scene images are taken into account. Otsu thresholding is employed for text binarization form simple background image and double edge features are employed for complex background text binarization. A framework for overlay text binarization is developed in[13]. Text polarity is determined for whether the light text with dark background or the dark text with light background. Then clustering is applied and binarization is achieved by exploiting Markov random field model. Finally, appropriate Binarized text is extracted using Log-Gabor filter.

The proposed text binarization approach is formulated based on three approaches; hill climbing algorithm, K-means clustering and graph cut algorithms. The descriptions of the employed approaches are discussed in this section briefly.

2. Segmentation Techniques

2.1 Hill Climbing Algorithm

Hill Climbing (HC) algorithms provides a unified framework for fast nonparametric optimization problem by determining peaks of cluster in colour histogram of an image automatically. In order to efficiently predict the peak value, where the histogram bins are exploited than the pixels themselves[14]. In order to find out the number of initial centroids (Peaks), the following steps are performed.

1. The colour histogram is given as an input to the hill climbing algorithm.
2. The peak selection process is started at a non-zero bin of the histogram and makes uphill moves until reaching a peak as follows:
   - The number of pixels of the current colour histogram bin is compared with the number of pixels of the neighbouring (left and right) bins.
   - If the neighbouring bins have different numbers of pixels, the algorithm makes an uphill move towards the neighbouring bin with larger number of pixels.
   - If the immediate neighbouring bins have the same numbers of pixels, the algorithm checks the next neighbouring bins, and so on, until two neighbouring bins with different numbers of pixels are found. Then, an uphill move is made towards the bin with larger number of pixels.
   - The uphill climbing is continued (repeat steps 2.1-2.3) until reaching a bin from where there is no possible uphill movement. That is the case when the neighbouring bins have smaller numbers of pixels than the current bin. Hence, the current bin is identified as a peak (local maximum).
3. Another unclimbed bin as a starting bin is selected and perform step 2 to find another peak. This step is

continued until all non-zero bins of the histogram are climbed (associated with a peak).

4. The identified peaks represent the initial number of clusters of the input image; thus these peaks are saved for further processing.

2.2 Graph Cut Segmentation

Graph cut segmentation is a global optimal key for image segmentation, where segmentation of an object from the background can be formulated as a binary labelling process. Given a set of labels L and a set of sites S, the labelling problem is to assign a label $f_p \in L$ to each of the sites $p \in S$.The graph cuts framework is presented[15] addresses the segmentation of a monochrome image, which solves a labelling problem with two labels. The label set is L = {0, 1}, where 0 corresponds to the background and 1 corresponds to the object. Let $f = \{ f_p | f_p \in L \}$ For a labelling, i.e. label assignments to all pixels. The formulated energy function is as follows:

$$E(f) = \sum_{p \in s} D_p(f_p) + \lambda \sum_{\{p,q\} \in N} \omega_{pq} . T(f_p \neq f_q) \quad (1)$$

On the right hand side of (1), the first term is a constraints from the observed data and measures how sites like the labels, that f assigns to them. Where D_p measures how well label f_p fits site p. $D_p(f_p)$is defined as the negative log likelihoods of the constructed foreground/background models. The second term is called the smoothness term and measures the extent to which f is not piecewise smooth where N is a 4-connected neighbourhood system or an 8-connected neighbourhood system. The smoothness term typically used for image segmentation is the Potts Model[16]. In image segmentation, the boundary lies on the edges in the image. A typical choice for ω *p,q* is,

$$\omega_{pq} = e^{\frac{(I_p - I_q)^2}{2\delta^2}} . \frac{1}{dist(p,q)} \quad (2)$$

Where and are the colour values of p and q and *dist*(*p,q*)is the distance between sites p and q. Parameter δ is related to the level of variation between neighbouring sites within the same object. The parameter λ is used to control the relative importance of the data term versus the smoothness term.

3. Proposed System

The proposed scene text binarization system is composed of five computational modules; automated peak selection, clustering analysis, image smoothness, graph cut and optimal cluster selection. The complete block diagram of the proposed system is shown in Figure 1.

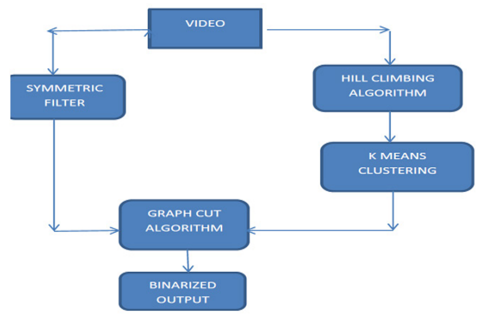

Figure 1. Block diagram of the proposed text binarization approach.

3.1 Clustering Analysis

Clustering is a major data mining task in dataset exploration and data partitioning. Among various clustering methods, simple iterative and prototype based K-means clustering technique is employed. However, for successful clustering result, the selection of number of initial centroid is most important. It is achieved by employing hill climbing approach. After selecting initial centroids similar characteristics (pixel values) are assigned to the nearby cancroids and each collection of pixel values are assigned to a group of cluster. Similarly the centroid of each cluster is updated until the number pixels in the cluster do not change. It is computationally an efficient process and then the clustered data points are given as an input to graph cut algorithm.

3.2 Symmetric Filtration

In order to mitigate edge effect of the given image, symmetric filter is employed to achieve effective representation of the given image. It smoothes and sharps the given scene image along with diagonal manner rather than rows and columns. Thus, the spatially varying parts are effectively utilized in this study which will improve the performance of the binarization process. They are fed to graph cut algorithm along with the clustered data points.

3.3 Graph Cut Segmentation

Generally, image segmentation algorithms involve assigning a label (such as disparity) to every pixel. There is a probability of sharpness discontinuities between the clustered data points. In order to overcome this, graph cut segmentation algorithm is employed that preserves sharp discontinuities such as object boundaries in the clustered data. To smooth the labelling process, symmetry filtered output and cluster image obtained from the clustering analysis is given as input to the graph cut segmentation module. Commonly, graph cut algorithm partitions directed or undirected graph into different group by two labelling approach. In order to provide energy minimized segmentation output, clustered and sharpened images are given as data cost and smoothness cost along with computed smoothness cost matrix to the graph cut module.

3.4 Optimal Cluster Selection

The optimal cluster selection is implemented to obtain the best clusters that has text region. First, the border clusters are discarded as they not related to text regions. The next Criterion to discard non text cluster uses the difference between the intensity means of successive clusters. It will remove the non text cluster between the text clusters. Finally, a validation measure based on the area of cluster is used to remove clusters having non text region. The remaining clusters are combined to form the final binarized image.

4. Experimental Results

The proposed approach is tested on videos taken from the You Tube Video Text (YVT) datasets in order to analyze the efficiency of the text binarization approach[17]. To improve the efficiency of the retrieval system, the salient features of CBIR and Text based retrieval system can be fused[18]. Figure 2 shows the localized and binarized output of input frame. Our algorithm is applicable to automated licence plate recognition system. Figure 3 shows the localized and binarized output of input frame. False Positive (FP), True Negative (TN), False Negative (FN), Sensitivity, Specificity and Accuracy.

(a) (b) (c)

Figure 2. (a) Input frame (b) Localized Output (c) Binarized output.

(a) (b) (c)

Figure 3. (a) Input (b) Localized Output (c) Binarized output.

FP is defined as non text pixels incorrectly identified as text pixels. TN is defined as non text pixels correctly identified as non text pixels. FN is defined as text pixels incorrectly identified as non-text pixels. The ability to identify text region is represented by sensitivity and it can be represented as follows:

$$\text{Sensitivity} = \frac{TP}{(TP+FN)} \tag{3}$$

The ability to identify non text region is represented by specificity of the system

$$\text{Specificity} = \frac{TN}{(FP+TN)} \tag{4}$$

$$\text{Precision} = \frac{TP}{(TP+FP)} \tag{5}$$

5. Conclusion

In this paper, an efficient text binarization system is implemented for indoor and outdoor scene text images. To facilitate the proposed framework, five computational modules are developed. At first, initial number of clusters is calculated by hill climbing algorithm followed by K-means clustering approach. Subsequently, the clusters are refined by graph cut algorithm that uses symmetric filtered image. Finally, optimal cluster selection selects the best clusters for text binarization. The proposed system is compared with two competing algorithms and the results shown the effectiveness of the system with improved accuracy. The proposed approach is tested on number of videos and the obtained results are good for text localization. Our algorithm is applicable to automated license plate recognition systems, especially to identify Indian vehicles license plate.

6. References

1. Cheriet M, Farrahi Moghaddam R, Hedjam R. A learning framework for the optimization and automation of document binarization methods. Computer Vision and Image Understanding. 2013; 117(3):269–80.
2. Badekas E, Papamarkos N. Optimal combination of document binarization techniques using a self-organizing map neural network. Engineering Applications of Artificial Intelligence. 2007; 20(1):11–24.
3. Mesquita RG, Silva RM, Mello CA, Miranda PB. Parameter tuning for document image binarization using a racing algorithm. Expert Systems with Applications. 2015; 42(5):2593–603.
4. Jacobs BA, Momoniat E. A novel approach to text binarization via a diffusion-based model. Applied Mathematics and Computation 2013; 225:446–60.
5. Lyu MR, Song J, Cai M. A comprehensive method for multilingual video text detection, localization and extraction. IEEE Transactions on Circuits and Systems for Video Technology. 2005; 15(2):243–55.
6. Wang B, Li XF, Liu F, Hu F Q. Color text image binarization based on binary texture analysis. Pattern Recognition Letters. 2007; 26(11):1650–7.
7. Su B, Lu S, Tan CL. Robust document image binarization technique for degraded document images. IEEE Transactions on Image Processing. 2013; 22(4):1408–17.
8. Naegel B, Wendling L. A document binarization method based on connected operators. Pattern Recognition Letters. 2010; 31(11):1251–9.
9. Gatos B, Pratikakis I, Kepene K, Perantonis SJ. Text detection in indoor/outdoor scene images. Proceedings First Workshop of Camera based Document Analysis and Recognition; 2005. p. 127–32.
10. Mancas-Thillou C, Gosselin, B. Color text extraction with selective metric based clustering. Computer Vision and Image Understanding 2007; 107(1):97–107.
11. Shi C, Xiao B, Wang C, Zhang Y. Adaptive graph cut based binarization of video text images. IEEE 10th IAPR International Workshop on Document Analysis Systems; 2012. p. 58–62.
12. Bai J, Feng B, Xu B. Binarization of natural scene text based on L1-Norm PCA. IEEE International Conference on Multimedia and Expo Workshops; 2013. p. 1–4.
13. Zhang Z, Wang W. A Novel Approach for Binarization of Overlay Text. IEEE International Conference on Systems, Man and Cybernetics; 2013 p. 4259–64.
14. Ohashi T. Fast segmentation of texture image regions based on hill-climbing. IEEE Pacific Rim Conference on Communications, Computers and signal Processing. 2003; 2:848–51.
15. Boykov Y, Jolly MP. Interactive graph cuts for optimal boundary and region segmentation. International Conference on Computer Vision. 2001; 1:105–12.
16. Boykov Y, Veksler O, Zabih R. Markov random fields with efficient approximations. IEEE Conference on Computer Vision and Pattern Recognition; 1998. p. 648–55.
17. Nguyen PX, Wang K, Belongie S. Video Text Detection and Recognition: Dataset and Benchmark; 2014. p. 776–83.
18. Jeyakumar V, Kanagaraj BR. Performance Evaluation of Image Retrieval System Based on Error Metrics. Indian Journal of Science and Technology; 2015. p. 117–21.