International Journal of Scientific Engineering and Research (IJSER)
ISSN (Online): 2347-3878
Impact Factor (2018): 5.426

Text Line Detection Using Connected Components

Safvana Sajeev[1], Ojus Thomas Lee[2]

[1]Post Graduation Student, College of Engineering Kidangoor

[2]Associate Professor & HOD, College of Engineering Kidangoor

Abstract: ***Today's digital era, the attention towards camera based text processing has increased many folds. This has led to the development of multiple text processing methods. Most of the procedures follow a scene text detection manner and further improvement is needed in camera captured images. In this research paper we present a survey and analysis of the existing methods in text line detection for camera captured images. In this work we have identified text-line detection methods such as MSER with Ada boost, State Estimation, Texture Based SVM, Chain Level Classification, CMSER Algorithm, Graph Model, and Fuzzy Method. MSER with Ada boost is the better one, it is robust and efficient.***

Keywords: Camera captured image processing, Connected Components, Character Segmentation, Classifier, Text Detection

1. Introduction

The importance of image processing has increased day by day due to the availability of cost effective devices for capturing photos and videos. In today's digital world, people captured photos and videos and pour into the social media. With the availability of digital cameras and smart phones, the camera based text detection has got an increased attention. Text is the most expansive means of communication. Text is embedded into documents or scenes for communication. It can be observable and readable by others. Several studies are conducted for text detection for camera captured images. The narrative in image contain valuable information and can be used for numerous image and video based applications like image search web site, information recovery, reviewing text and identify text in mobile devices. The size, font, style of text and the back ground with snow, rock make text detection in camera captured images more difficult. Several methods have been found for find the solution to this problem.

Detecting text in constrained, controlled environments can typically be accomplished by using heuristic-based approaches, like exploiting gradient information or the fact that text is typically grouped into paragraphs and characters appear on a straight line. The text detection methods are generally connected components or texture based. The connected components method segment the character's into smaller components. Arrange connected components to text/background by using their geometrical properties. The connected components based method is difficult to run when the text contain noise. The connected components method is most efficient for extracting the text from camera captured images. Texture based images are used here to extract textual properties of the images. Neural networks can be used to differentiate text from the document images.

Give an image Maximal Stable Extremal Region (MSER) tries to find connected (and nearly connected) regions. Here need to take the parameters such as the illumination, orientation, size etc... In gray scale images, the local variations are analyzed to locate the regions. The combination of texture based method and connected components extract the wavelets features of image and then classify using neural network. The texture based methods perform well in noisy image. The process is usually slow. Text detection procedures are categories into two:

1) *Connected Components (CC):* Using Maximally Stable Extremal Region Algorithm (MSER) the images are extracted. The extracted components are then classified by using SVM/Ada Boost classifier.
2) *Region Based (RB):* Consider text has different texture features compared to the background. In pixel values the neighboring gray values are considered as features. They are classified by using SVM to differentiate the text/non-text content.

Text detection process is the major step in text line detection task. It is used in Optical Character Recognition (OCR) and preprocessing algorithms. Text line detection is difficult in curved text and blurred text. The English language text in the images will be captured.

The main text detection challenges are divided into three groups:

1) *Variety of natural images:* The characters in natural images are in different font style, different size, unique color, and unique font alignment.
2) *Problems in backgrounds:* The background includes grasses bricks, rocks, are leads to more complexity for identifying text.
3) *Interference factors:* The main inference factors are blurring.

To solve these challenges a robust, efficient mechanism for text line detection is needed. The main contributions of the work include study on

1) MSER with Ada boost
2) State Estimation
3) Texture Based SVM
4) Chain Level Classification
5) CMSER Algorithm
6) Graph Model
7) Fuzzy Method

The rest of this paper is organized as follows. Different methodologies for text detection are reviewed in Section 2. Final remarks are presented in Section 3.

2. Methodologies for Text Detection

Text plays a major role in our day to day life. With arrival of social media and smart phones, taking photos and sharing them has become common. Frequently, these shared images become useful for people in crime detection, evidence collection and many similar activities. The text in the images play important role in these activities. Hence text detection in images has captured attention these days. Text detection means to identify text in a given input images. In the recent past, researchers have proposed different techniques to extract characters from images. Text is embedded in image either scene text or caption text. In scene text the text appears in an image that was captured by a camera in an outdoor environment.

Text detection in natural image is carried out by constructing a bounded rectangle box around the text. There are many techniques existing for the detection of text in an image. Capturing text from indoor and outdoor images is difficult due to variety of backgrounds.

Optical Character Recognition (OCR) was the technology used to detect text from scanned images and printed images. But it faces a lot of difficulties like complex backgrounds, diversity of scene text and interference factor like, font, orientation of text. So to overcome the above mentioned difficulties, researchers have developed techniques like Ada boost, SVM, Filters, and Convolution Neural Networks (CNN). Text detection can have variety of vision based application such as robot navigation, geo-location, industrial automation, number plate recognition. This paper gives a detailed survey of existing techniques for text detection.

After reviewing the literature in the existing domain we have found that there exists the following methods for solving the problem.
1) MSER with Ada boost
2) State Estimation
3) Texture Based SVM
4) Chain Level Classification
5) CMSER Algorithm
6) Graph Model
7) Fuzzy Method

2.1 MSER with Ada boost

Smart phones, laptops have become popular text extraction and identification platform for image pre possessing. With the increased availability of smart phones and digital cameras in the society, has given a thrust to development of application that uses photos. These applications are useful for visually impaired peoples, translators for tourists, information retrieval systems and automatic robot navigation system. The scene text images affect photo metric degradation and geometrical distortions.

[1], [2], [3] Propose a connected component procedures that extraction the connected component and form a group of individual components that contain text.

The procedure has certain challenges:
1) Identification of a bit of text.
2) Detect characters similar to connected components.
3) Extraction of non-text connected components.
4) Deduce character groups in connected components.

Procedure contains mainly 3 phases. They are, character generation, character normalization, and refine.

1) Character Generation (CG): The phase use connected component-based procedure, and concentrate challenge 4 initially. Challenge 2 and 3 are used to identify the connected component from images by using MSER techniques. Recognized components are segmented into clusters. For finding the relationship between these components authors train an Ada Boost classifier that finds adjacency relationship between the clusters.

2) Character Normalization (CN): Standardize the characters identified in phase I.

3) Refine(R): Here decide the image contain text content or not.

A major problem with this process is that, the process become difficult due to the scale, skew, and color of each character.

[4], [5] are suggest procedure to detecting character content from natural images. This procedure contains 4 steps:
a) *Character Candidate Generation (CCG):* By using MSER the character candidates are generated.
b) *Character Candidate Elimination (CCE):* Here pruning occur. The tree data structure is used. Elimination can be done using two methods,
   - Linear Reduction
   - Tree Accumulation.
c) *Character Candidate Normalization (CCN):* Normalize the word region.
d) *Non- text Filtering (NF):* Here decide the image contain character/not.

Gap Identified: The main issue with MSER based methods is the difficulty in detecting blurred images.

Figure 1: MSER with Ada boost

2.2 State Estimation

Character recognition and extraction is a major phase in OCR and number plate recognition. It is also used for camera captured document images to detect or extract texts and characters. The method is apply both camera captured and handwritten documents. [6], [7] Initiate state estimation procedure for the extraction of text in natural images. Here algorithm has two phases. First phase using the connected components, interline spacing is calculated. The scale selection is an essential step that is used to detect features from unknown data. State estimation is an essential for artificial document image processing. For example, consider

two connected components, and find out whether these components are in a same word or not. The result may be ambiguous unless we know their states. The next phase of the procedure identifies the text blocks and extracts text lines using the estimated states. The text line extraction procedure use bottom-up grouping. The system works as follows, initially take the input images and the using the state estimation the state information of the images are taken. Then identifies the text blocks and extract the characters which are grouped by using clustering or bottom up grouping. Later by using the states the text lines are identified. [8] Suggest a mechanism to identify the characters from natural images. Previous procedures for character identification are divided into two categories.

1) Regional Based: Attempt to detect text regions by using MSER.
2) Connected Component Based: Component extraction is done. Using SVM/Ada Boost classifier, the components that contain characters are identified and filter out the rest.

The techniques for detecting scene text are robust by incorporating details of regions into a connected component. Characters are identified using the region detector. The region detector analyzes each layer/pixel of the image. The texts are linked via minimum spanning tree. At the end of the process the state estimation tool is used to group the texts components.

Gap Identified: The main issue of this method is, it can only train the Chinese and English texts.

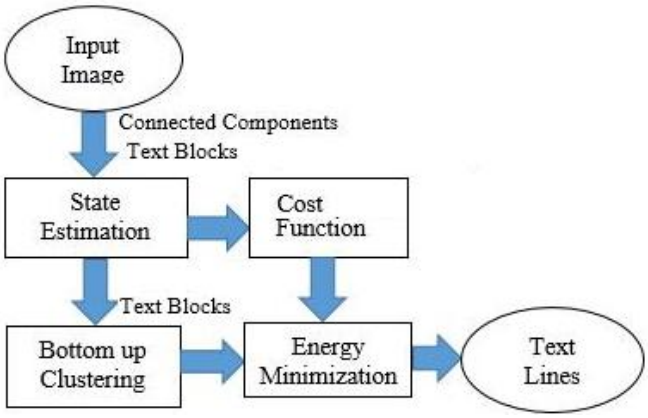

Figure 2: State Estimation

2.3 Texture Based SVM

Image texture look like ordered values. The values can be calculated for finding the texture features of image. The texture values give details of the image. The texture can be created manually or it is taken when the image is captured. Support Vector Machine (SVM) provides a detection method for reducing texture values.

[9], [10], [11] Proposed texture based SVM method. Texture classification is like score image. Each pixel is denoted as the possibility of the part of region which contain texts. The texts score is repeatedly applied to the continuously adaptive mean shift algorithm (CAMSHIFT). Text regions are identified by applying a CAMSHIFT to the results of the texture analysis. The CAMSHIFT and SVM produce robust and efficient detection. The texts are extracted by apply the CAMSHIFT algorithm to the texture score image. The system analyzes only a few part of the input image. The system confined moderate texture analysis for less pertinent pixels.

[12] Propose a connected component with SVM method to identify the characters from the image. Detecting texts from an image is a complex issue. It attracts many researchers due to vast availability and easiness of handheld devices such as smart phones and digital cameras. The task is still the challenging one because of the varying properties of texts like font, color, size of the character, illumination and orientation of character. The connected component (CC) based approach segment the image into regions, select the appropriate regions and locate the text from those detected regions. An image is subdivided into smaller components and they are grouped together based on their geometric properties. The approach uses the edge detection method. To detect the textual components and classify them as textual regions, mostly a geometrical analysis is required as the CC-based approach is based on geometrical features. The approach has minimum computational complexity due to several segmented components.

Gaps Identified: Difficult to classify very small text with low contrast.

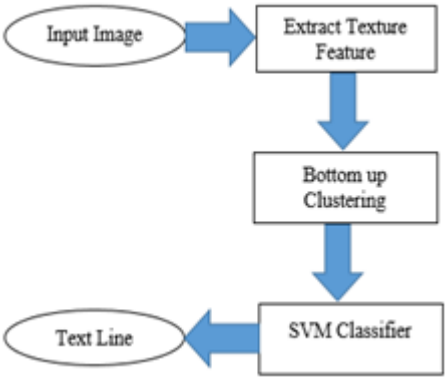

Figure 3: Texture Based SVM

2.4 Chain Level Classifier

Identify texts in an arbitrary orientation is a challenging task and has received less attention. [13] Propose arbitrary text detection in natural images. Chain level classifier is a system that produces robust result for detecting text. System is fast, efficient and works under various condition. Whenever the system is applied directly to detect arbitrary texts then, the conventional rules for horizontal texts become false positive. To overcome the false positive here a low level features are used. System efficiently detects the arbitrary oriented texts.

System contains 4 parts:
1) *Component Extraction (CE):* Here perform Edge detection and extract into components.
2) *Component Analysis (CMPA):* Several components are identified at these stages. Here analyzes the components and choose the filter to identify the non-texts components.

3) *Candidate Linking (CL):* The characters are linked pair wise. The adjacent characters are linked by taking their geometric similarity.
4) *Chain Analysis (CA):* Analyze the pair of characters and determine the texts using a classifier.

Gaps Identified: The system produce wrong result when the Segmentation becomes hard. System needs further iteration.

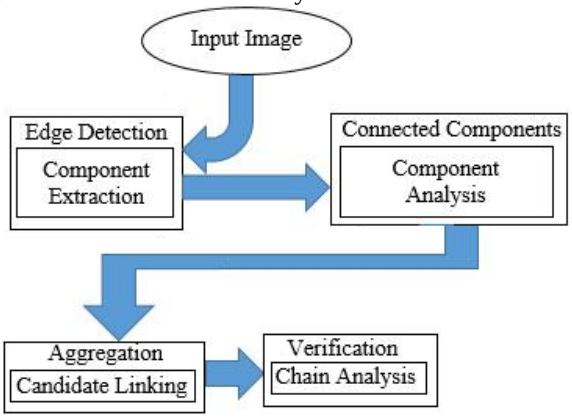

Figure 4: Chain Level Classifier

2.5 CMSER Algorithm

In our daily life we all see sign boards in high ways. The sign boards become an important part of our life. The sign board images contain written words, texts, arrow marks, numbers and symbol. These sign boards help the driver to know which road to be chosen. An HD front camera in the car captures the sign board images. Such an application will be helpful for blind persons for identifying land marks and directions.

[14] Suggests a method to detect road sign text. First step is to take the sign board image using HD front camera. The MSER procedure detects the characters from the sign board image. Eliminate the non-text ones using heuristic rules. The final step is to determine the likeness of characters based on their orientation/geometric property and group them. [15], [16] Proposed an identification mechanism for traffic sign boards. A number of applications are there like driver assistance system, autonomous vehicle. Several mechanisms are used to detect texts in road sign boards like,

1) Maximally Stable Extremal Region (MSER): The algorithm includes observing the candidates for traffic signs among the outlined scene search regions. MSERs is a region detector, it is robust and detect high contrast regions.

2) Histogram of Oriented Gradient (HOG): HOG is a widely used shape descriptor used to detect traffic sign. To extract descriptors of an image the HOG divide the image into blocks which are formulated by using overlapping cells. Each pixel formulates a histogram that contains magnitude and orientation. To extract color of an image, the HOG calculate color descent of color channels and HOG is the most popular color detection method used.

3) Optical Character Recognition (OCR): OCR electronically extracts the characters from the image. Here recognition of printed or written text characters are done by using a computer. The optical character reader is a mechanical or electronic conversion device. Text scanning occurs character by character.

After comparing different techniques used for the detection and recognition of traffic signs, MSER and OCR are the most popular and efficient method. OCR provides better recognition performance.

Gaps Identified: Sometimes text is not identified so, more efficient mechanism is needed for detection and recognition from traffic signs.

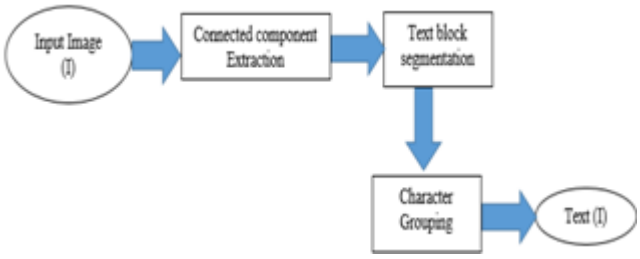

Figure 5: CMSER Algorithm

2.6 Graph Model

The increased advancement of digital images leads to development a fast image analysis technology. The scene text detection mechanism affects difficulty when the segmentation procedure occurs. [17] Proposed an efficient component based text detection technique that incorporate the region approach. The text information in the image is shown as a graph which contains nodes as characters. Scene text approach uses the graph model. Here MSER technique is used divide the image into sub regions. Construct a graph that contains nodes as region where nodes are representing characters. MSER efficiently labels the text or non-texts ones and reduces the cost using graph cut algorithm. The classifiers group the similar characters and form a word, since most non-text MSERs are eliminated, by using a trained classifier.

Steps involved in the process are listed below.
1) Initially MSER extracts the image into several sub regions.
2) Construct graph that contain nodes as region.
3) Label the text and non-text ones.
4) The labeling result is minimizing cost using a graph cut algorithm.

[18], [19] Proposed techniques for identifying curved texts from an image scene. Initially components are extracted by using MSER. In component approach the segmentation phase becomes difficult. After the segment, components are classified based on their similarity. Then each component is binarized and the text regions are detected. Here make two assumptions,
1) Characters belong to the English alphabets
2) Curved texts are considered

An algorithm in to do this contain mainly 3 steps.
1) Highlight the important information using segmentation.

2) Apply classification procedure for separate the character and non-character.
3) Group the characters based on their similarity and form a word.

The procedure is common universal since it makes only very little assumption about features of the text. The component deletion and insertion is easy for specific purpose for example, it is easy to group horizontal and vertical texts.

Gaps Identified: Some texts do not get identified in MSER, because of the illumination. When the assumption is fails the false positive may be detected.

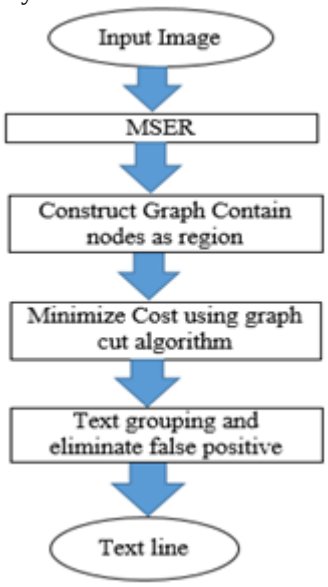

Figure 6: Graph Model

2.7 Fuzzy Method

Identification and extraction of text contents in videos are essential for information retrieval purpose. Though it is useful for the society, it is a challenging task. [20], [21] Produce a fuzzy ensemble technique for the detection of text in video. The advantage of fuzzy method is that it provides the temporal data regarding the static text in the video. The system initially extracts the features of an image by taking video frames. The method contains 3 steps. They are,

1) A fuzzy clustering algorithm is applied to the individual frames. The advantage of fuzzy clustering is more flexible during the time of clustering.
2) Apply clustering ensemble to the clusters and then the result integrate with the result of fuzzy C means algorithm in individual frames.
3) Determine the correct cluster that contains texts.

Gaps Identified: The application of the clustering ensemble with different features leads to further iteration of the system.

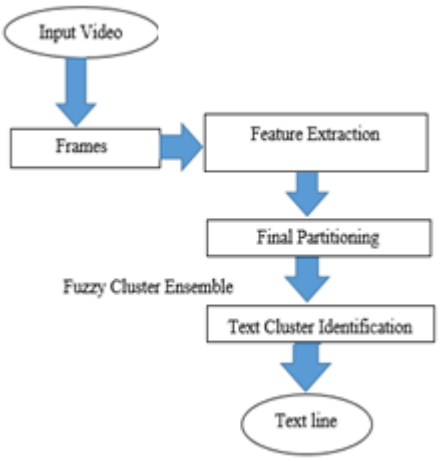

Figure 7: Fuzzy Method

Table 1: Matrices

Method	*Precision*	*Recall*	*F-Score*
MSER with Ada boost	71.00%	61.00%	66.00%
Texture Based SVM	71.00%	68.00%	69.00%
Chain Level Classifier	68.00%	66.00%	66.00%
Graph Model	83.30%	63.10%	71.00%
CMSER Algorithm	79.00%	49.00%	53.00%
Fuzzy Method	96.71%	92.04%	94.00%

3. Conclusion

The survey is based on text detection using image processing technique. Different types of difficulties have been faced when identifying text from images and videos. To overcome these difficulties, different techniques has been introduced like MSER with Ada boost, State Estimation, Texture Based SVM, Chain Level Classification, CMSER Algorithm, Graph Model and Fuzzy Method. OCR, SVM, Ada boost are different types of classifiers used for detecting the text content. MSER with Ada boost is the better one, it is robust and efficient.

References

[1] H. I. Koo, D. H. Kim, Scene text detection via connected component clustering and nontext filtering, IEEE transactions on image processing 22 (6) (2013) 2296–2305.
[2] X. Chen, A. L. Yuille, Detecting and reading text in natural scenes, in: Computer Vision and Pattern Recognition, 2004, CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on Vol.2, IEEE, 2004, pp, II–II.
[3] M. S, detecting text in natural scenes with connected components clustering and non text filtering, in: International journal of Engg and technology (IRJET2016), IRJET, 2016, pp. 725–728.
[4] X.-C. Yin, X. Yin, K. Huang, H.-W. Hao, Robust text detection in natural scene images, IEEE transactions on pattern analysis and machine intelligence 36 (5) (2014) 970–983.
[5] X. Yin, X.-C. Yin, H.-W. Hao, K. Iqbal, Effective text localization in natural scene images with mser, geometry-based grouping and adaboost, in: Proceedings

of the 21st International Conference on Pattern Recognition (ICPR2012), IEEE, 2012, pp. 725–728.
[6] H. I. Koo, N. I. Cho, State estimation in a document image and its application in text block identification and text line extraction, in: European Conference on Computer Vision, Springer, 2010, pp. 421–434.
[7] H. I. Koo, N. I. Cho, Text-line extraction in handwritten chinese documents based on an energy minimization framework, IEEE Transactions on Image Processing 21 (3) (2012) 1169–1175.
[8] Y.-F. Pan, X. Hou, C.-L. Liu, et al., A hybrid approach to detect and localize texts in natural scene images, IEEE Transactions on Image Processing 20 (3) (2011) 800.
[9] K. I. Kim, K. Jung, J. H. Kim, Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm, IEEE Transactions on Pattern Analysis and Machine Intelligence 25 (12) (2003) 1631–1639.\
[10] K. Jung, K. I. Kim, A. K. Jain, Text information extraction in images and video: a survey, Pattern recognition 37 (5) (2004) 977–997.
[11] L. G´omez, D. Karatzas, Mser-based real-time text detection and tracking, in: Pattern Recognition (ICPR), 2014 22nd International Conference on, IEEE, 2014, pp. 3110–3115.
[12] S. Unar, A. Hussain, M. Shaikh, K. H. Memon, M. A. Ansari, Z. Memon, A study on text detection and localization techniques for natural scene images, IJCSNS 18 (1) (2018) 100.
[13] C. Yao, X. Bai, W. Liu, Y. Ma, Z. Tu, Detecting texts of arbitrary orientations in natural images, in: 2012 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2012, pp. 1083–1090.
[14] M. S. Hossain, A. F. Alwan, M. Pervin, Road sign text detection using contrast intensify maximally stable extremal regions, in: 2018 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE), IEEE, 2018, pp. 321–325.
[15] K. Sumi, M. Arun Kumar, Detection, and recognition of road traffic signs-a survey, International Journal of Computer Application (0975-8887) 160 (3).
[16] Y. Tang, X. Wu, Scene text detection and segmentation based on cascaded convolution neural networks, IEEE Transactions on Image Processing 26 (3) (2017) 1509–1520.
[17] C. Shi, C. Wang, B. Xiao, Y. Zhang, S. Gao, Scene text detection using graph model built upon maximally stable extremal regions, Pattern recognition letters 34 (2) (2013) 107–116.
[18] J. Fabrizio, M. Robert-Seidowsky, S. Dubuisson, S. Calarasanu, R. Boissel, Textcatcher: a method to detect curved and challenging text in natural scenes, International Journal on Document Analysis and Recognition (IJDAR) 19 (2) (2016) 99–117.
[19] D. He, X. Yang, W. Huang, Z. Zhou, D. Kifer, C. L. Giles, Aggregating local context for accurate scene text detection, in: Asian Conference on Computer Vision, Springer, 2016, pp. 280–296.
[20] J. Gllavata, E. Qeli, B. Freisleben, Detecting text in videos using fuzzy clustering ensembles, in: Multimedia, 2006. ISM'06. Eighth IEEE International Symposium on, IEEE, 2006, pp. 283–290.
[21] K. O. Aziz, Better text detection through improved k-means-based feature learning.

Author Profile

Safvana Sajeev, She is a post-graduation student in College of Engineering Kidangoor. Specialization in Computer and Information Science. She received the graduation in Computer Science and Engineering from Gurudeva Institute of Science and Technology. Her areas of interest are Image Processing and Databases.

Ojus Thomas Lee, is an Associate Professor in College of Engineering Kidangoor. His areas of interest are Databases, Cloud Computing, Sensor Networks and Image Processing.