International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 04 Issue: 04 | Apr -2017 www.irjet.net p-ISSN: 2395-0072

A Novel Text Detection System

[1]Prof. Harish Barapatre, [2]Prashant Athavale, [3]Sameer Suryavanshi,
[4]Abhishek Deware
Yadavrao Tasgaonkar Institute Of Engineering and Technology Dept. Of Computer Engineering

---

Abstract - *The proposed method is a novel method which is using three new features to detect text objects comprising two or more isolated characters in images. Each character is a part in the model and every two neighboring characters are connected by a link. Two characters and the link connecting them are defined as a text unit. For every candidate part we compute character energy and link energy character stroke forms two edges with high similarities in length, curvature, and orientation and the similarities in color, size, stroke width, and spacing between characters. we combine character and link energies to compute text unit energy which measures the likelihood that the candidate is a text object. Our proposed system can inherits properties of characters and discriminate text from other objects effectively.*

1. INTRODUCTION

Identification of text in an image is almost effortless for human being but it is very difficult for machine. Detecting text in natural images, as opposed to scans of printed pages, faxes and business cards, is an important step for a number of Computer Vision applications, such as computerized aid for visually impaired, automatic geo coding of businesses, and robotic navigation in urban environments.

Visual saliency is fundamental to the human visual system, and there is need to process it. As such it has been a well studied problem within multiple disciplines, including cognitive psychology, neurobiology, and computer vision. The aim of salient object detection is to highlight the whole attention grabbing object with well-defined boundary. Previous saliency detection approaches can be broadly classified into either local or global methods.

Measures of 'objectness' have built upon the saliency detection in order to identify windows within an image that are likely to contain an object of interest. Saliency detection models facilitate scene text detection; they share a common inherent limitation, which is that they are distracted by other salient objects in the scene. The approach here differs from these existing methods in that there is proposed a text-specific saliency detection model (*i.e.* a characterless model) and demonstrate its robustness when applied to scene text detection.

This text detection is capable of identifying individual, bounded units of text, rather than areas with text-like characteristics. The unit in the case of text is the character, and much like the 'object', it has a particular set of characteristics, including a closed boundary. Text is made up of a set of interrelated characters. Therefore, effective text detection should be able to compensate for, and exploit these dependencies between characters.

The three new characterness cues developed, instead of simple linear combination, a Bayesian approach is used to model the joint probability that a candidate region represents a character. The probability distribution of cues on both characters and non-characters are obtained from training samples.

In order to model and exploit the inter-dependencies between characters we use the graph cuts algorithm to carry out inference over an MRF designed for the purpose.

This approach is first to present a saliency detection model which measures the characterness of image regions. This text-specific saliency detection model is less likely to be distracted by other objects which are usually considered as salient in general saliency detection models. Promising experimental results on benchmark datasets demonstrate that characterness approach outperforms the state of-the-art.

1.1 EXISTING SYSTEM

In Previous segmentation methods, the process applies only for segmenting an object and that too only for single object. Most of the methods applies only for binary or gray images. In existing retrieval system, various algorithms proposed to improve the Retrieval properties between two images. They are such as BoW (Bag of words), GIST detectors, MSER detectors, GMM and SIFT are used to extract the feature. Among this, mixture models are commonly used in all detectors. Each model improves at least any of the parameter which improves the matching property. Pixels from the segmented object is known as feature pixels. Depend on features of pixels, the pixel matching performed between multiple images and then the image retrieved .The datasets used are commonly availed Real world datasets such as Oxford buildings and INRIA holidays.

A Maximally Stable Extremal Region (MSER) is a connected component of an appropriately thresholded image. The word 'extremal' refers to the propertythat all pixels inside the MSER have either higher(bright extremal regions) or lower (dark extremalregions) intensity than all the pixels on its outerboundary. The 'maximally stable' in MSER describes the property optimized in the threshold selection

process. The set of extremal regions E, i.e., the set of all connected components obtained by thresholding, has a number of desirable properties. Firstly, a monotonic change of image intensities leaves E unchanged, since it depends only on the ordering of pixel intensities which is preserved under monotonic transformation. This ensures that common photometric changes modeled locally as linear or affine leave E unaffected, even if the camera is non-linear (gamma-corrected).Secondly, continuous geometric transformations preserve topology–pixels from a single connected component are transformed to a single connected component

1.2 DISADVANTAGES

- Most of the process concentrates either Segmentation or Retrieval.
- Here the image can be described as the histogram which is used to compute a similarity between the pair of images.
- By using K-mean clustering algorithm the vocabularies are build. But its results usually scale poorly with the size of the vocabulary.

2. LITERATURE REVIEW

Existing scene text detection approaches generally fall into one of three categories, namely, texture-based approaches, region-based approaches, and hybrid approaches.

Texture-based approaches extract distinct texture properties from sliding windows, and use a classifier to detect individual instances. Some widely used features include Histograms of Gradients (HOGs), Local Binary Patterns (LBP), Gabor filters and wavelets. Foreground regions on various scales are then merged to generate final text regions.

Region-based approaches on the other hand, first extract potential characters through edge detection, color clustering or Maximally Stable Extremal Region (MSER) detection. After that, low level features based on geometric and shape constraints are used to reject non-characters. As a final step, remaining regions are clustered into lines through measuring the similarities between them.

A typical example of region-based approaches was a local image operator, called Stroke Width Transform (SWT), assigns each pixel with the most likely stroke width value, followed by a series of rules in order to remove non-characters.

Hybrid approaches are a combination of texture based and region-based approaches. Usually, the initial step is to extract potential characters, which is the same as region based approaches. Instead of utilizing low level cues, these methods extract features from regions and exploit classifiers to decide whether a particular region contains text or not, and this is considered as a texture-based step.

1. Itti [et al., 1998] has said that a saliency map used to find regions of interest. Saliency map is used to represent the saliency at every location in the visual field. Multi scale image features such as color, intensity and orientation are combined into a single topographical saliency map. A dynamic neural network then selects attended locations in order of decreasing saliency. This system breaks down the complex problem of scene understanding by rapidly selecting, in a computationally efficient manner.

It detects a target which differs from surrounding distracters by its unique size, intensity, color or orientation. The efficiency of this approach for target detection critically depends on the features types implemented.

2. Matas [et al., 2002] has explained MSER; the regions are defined solely by an extremal property of the intensity function in the region and on its outer boundary. MSER is a method for blob detection in images. The MSER algorithm extracts from an image a number of co-variant regions, called MSERs. MSER is a stable connected component of some gray-level sets of the image. Extremal regions possess highly desirable properties:

The set is closed under

1. Continuous transformation of image coordinates

2. Monotonic transformation of image intensities.

An efficient and practically fast detection algorithm is presented for an affinely-invariant stable subset of extremal regions, the maximally stable extremal regions (MSER). This operation can be performed by first sorting all pixels by gray value and then incrementally adding pixels to each connected component as the threshold is changed. The area is monitored. Regions such that their variation with respect to the threshold is minimal are defined maximally stable.

MSER doesn't preserve sharp edges. It has some blurring effect, so it doesn't provide accurate accuracy and precision.

3. Alexe [et al., 2010] stated that objectness, built upon the saliency detection in order to identify windows within an image that are likely to contain an object. He has presented a generic objectness measure, quantifying how likely it is for an image window to contain an object of any class. They have explicitly trained it to distinguish objects with a well-defined boundary in space. The measure combines in a Bayesian framework several image cues like Color Contrast(CC), Edge Density (ED) and Super pixels Straddling (SS), measuring characteristics of objects, such as appearing different from their surroundings and having a closed boundary. This includes an innovative cue measuring the closed boundary characteristic, the combined measure to perform better than any cue alone.

This finds out salient object from surrounding, but is not that much efficient for character or text detection. It requires character features related cues to compute the characterness score.

4. Epshtein [et al., 2010] presented a novel image operator that seeks to find the value of stroke width for each image pixel, and demonstrate its use on the task of text detection in natural images. The suggested operator is local and data dependent, which makes it fast and robust enough to eliminate the need for multi-scale computation or scanning windows. Its simplicity allows the algorithm to detect texts in many fonts and languages. The SWT converts the image from containing gray values to an array containing likely stroke widths for each pixel. This information suffices for extracting the text by measuring the width variance in each component because text tends to maintain fixed stroke width. This puts it apart from other image elements such as foliage and detects the text.

5. Kumar [et al., 2010] proposed method of robust extraction of text from scene images Scene images usually have non-uniform illumination, complex background and existence of text-like objects. Due to non-uniform illumination the contrast of text pixels sinks and merges with background. As a result the common assumption of a homogeneous text region on a nearly uniform background cannot be maintained in real application. Text extraction method, which can utilize the available contrast of text pixels. Initially the colours of input images were reduced to small numbers using quantization method. Next, colour reduced images were converted to gray scale and pixels intensity values were estimated. Based on the intensity values, the pixels were separated into three layers of connected components. Missing texts were recovered using image binarization technique. Then the geometrical features of the entire connected components in each layer were analyzed to extract the text regions from non text regions. And finally layers were merged together and using a Sobel edge projection profile the text line was verified.

6. Shahab [et al., 2012] compared the performance of four different saliency detection models at scene text detection. These models are Itti's model, Harel's Graph Based Visual Saliency Model, Torralba's Model and Zhang's Fast Saliency Model.They have evaluated these four state-of-the-art models of visual attention for the task of scene text detection. They evaluated that which models of visual attention are best suited for the task of text detection in natural scenes. The experimental results showed that Torralba's model performed best for separation of text elements from non-text elements (background). They also identified the parameter combination for Itti's method which can be used for the task of text detection.

7. Li & Shen [2013] adopted model using edge preserving MSER to extract text region and then minimization of an energy function yields a binary label for image region which indicates that region is character or not. Finding text in natural images has been a challenging task in vision. At the core of state-of-the-art scene text detection algorithms are a set of text-specific features within extracted regions. In this paper, they have attempted to solve this problem from a different prospective. They have showed that characters and non-character interferences are separable by leveraging the surrounding context. Surrounding context, in work, is composed of two components which are computed in an information-theoretic fashion. Minimization of an energy cost function yields a binary label for each region, which indicates the category it belongs to. The proposed algorithm is fast, discriminative and tolerant to character variations and involves minimal parameter tuning.

8. Yin [2014] proposed model of extracting region by MSER then text candidate construction by single link clustering. After that character classifier estimates non texts. Text classifier is trained to decide whether text candidate corresponding to the true text or not. Text detection in natural scene images is an important prerequisite for many content-based image analysis tasks. In this paper, they proposed an accurate and robust method for detecting texts in natural scene images. A fast and effective pruning algorithm is designed to extract

Maximally Stable Extreme Regions (MSERs) as character candidates using the strategy of minimizing regularized variations. Character candidates are grouped into text candidates by the single-link clustering algorithm, where distance weights and clustering threshold are learned automatically by a novel self-training distance metric learning algorithm. The posterior probabilities of text candidates corresponding to non-text are estimated with a character classifier; text candidates with high non-text probabilities are eliminated and texts are identified with a text classifier. The proposed system is evaluated on the ICDAR 2011 Robust Reading Competition database; the f-measure is over 76%, much better than the state-of-the-art performance of 71%. Experiments on multilingual, street view, multi-orientation and even born-digital databases also demonstrate the effectiveness of the proposed method.

9. Ryu [2014] Explained method based on extracting connected components and group them into text lines. Firstly, CC extraction method applied and then text line segmentation algorithm is used. Text-line extraction in handwritten documents is an important step for document image understanding, and a number of algorithms have been proposed to address this problem. However, most of them exploit features of specific languages and work only for a given language. In order to overcome this limitation, they have developed a language-independent text-line extraction algorithm. Our method is based on connected components (CCs), however, unlike conventional methods; they have analyzed strokes and partition under-segmented CCs into normalized ones. Due to this normalization, the proposed method is able to estimate the states of CCs for a range of different languages and writing styles. From the estimated states, they built a cost function whose minimization yields text-lines. Experimental results show that the proposed method yields the state-of-the-art performance on Latin-based and Chinese script databases.

Thus we have seen various existing systems which are used to detect the text from images. But there is not any system which gives accurate and precise output. So for that purpose, there is necessity to implement a new system which should be based on character features. eMSER [edge preserving Maximally Stable Extremal Region] algorithm is based on region based approach. Three novel cues are then computed, each of which independently models the probability of the region forming a character. These cues are then fused in a Bayesian framework, where naïve Bayes is used to model the join probability. The posterior probability reflects the "characterless" if the corresponding image patch.

3. PROPOSED SYSTEM

The proposed scene text detection is divided into various parts, having characterness model and labeling and grouping. Specifically, in characterness model, in which perceptually homogeneous regions are extracted by a modified MSER-based region detector. Three novel characterness cues such as Stroke Width (SW), Perceptual Divergence (PD) and Histogram of Gradient at Edges (eHOG) are then computed, each of which independently models the probability of the region forming a character. These cues are then fused in a Bayesian framework, where Naive Bayes is used to model the joint probability. The posterior probability reflects the 'characterness' of the corresponding image patch. In order to consolidate the characterness responses, there is designing of a character labeling method, which separates characters from non characters. Minimized by graph cuts is used to combine evidence from multiple per-patch characterness estimates into evidence for a single character or compact group of characters. Finally, verified characters are grouped to readable text lines via a clustering scheme, containing two normalized features such as characteristic scale and major orientation.

Two phases of experiments are conducted separately in order to evaluate the characterness model and scene text detection approach as a whole. In the first phase, compare the proposed characterness model with ten state of-the-art saliency detection algorithms on the characterness evaluation task. For this, there are used dataset ICDAR 2013, which consists of 229 images. Randomly 100 images of this dataset are used here for evaluation. In the second phase, use bounding boxes of detected text lines to compare against state of- the-art scene text detection approaches. For this purpose, there are used 2 datasets, such as ICDAR 2003 and ICDAR 2011 dataset.

4. METHODOLOGY

4.1 BLOCK DIAGRAM

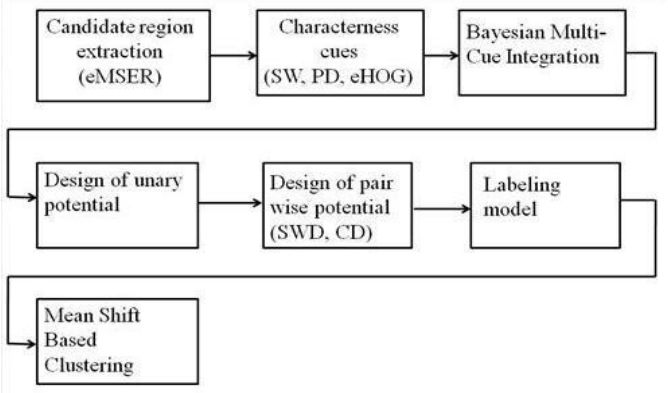

Fig1 : Block diagram

4.2 The Characterness Model

The characterness model computes the characterness score of the text in the image. Candidate region extraction via eMSER, calculating characterness cues (Stroke width, Perceptual divergence and Histogram of gradient) and finally calculating posterior probability of the region that region is character is come under the characterness model. This indicates that how character is salient from the background.

4.2.1 Candidate Region Extraction

MSER is an effective region detector which has been applied in various vision tasks, such as tracking, image matching, and scene text detection amongst others. The MSER detector is thus particularly well suited for identifying regions with almost uniform intensity surrounded by contrasting background. For the task of scene text detection, although the original MSER algorithm is able to detect characters in most cases, there are some characters

Fig. 2 . MSERvs eMSER.Cases that the original MSER fails to extract the characters while modified eMSER succeeds. (a) Original text. (b) Original MSER. (c) eMSER

This tends to degrade the performance in the scene text detection algorithms. To address this problem, there is used eMSER algorithm.

4.2.2 Characterness Cues

Characters attract human attention because their appearance differs from that of their surroundings. Three novel cues to measure the unique properties of characters.

4.2.3 Stroke Width (SW): Stroke width has been a widely exploited feature for text detection. In particular, SWT computes the length of a straight line between two edge pixels in the perpendicular direction, whichis used as a preprocessing step for later, a stroke is defined as a connected image region with uniform color and half-closed boundary. Although this assumption is not supported by some uncommon type faces, stroke width remains a valuable cue. The stroke width cue of region r is defined as:

$$SW(r) = Var(l) / E(l)^2 \quad \ldots(1)$$

Where,

$E(l)$ – stroke width mean

$Var(l)$ - stroke width variance

In Fig.3.4., we use color to visualize the stroke width of exemplar characters and non characters, where larger color variation indicates larger stroke width variance and *vice versa*. It shows that characters usually have small SW value.

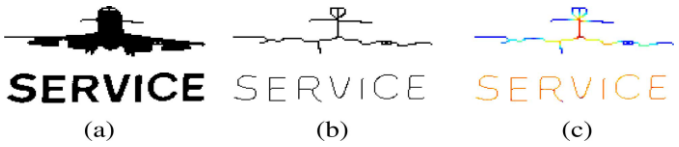

Fig.3 .Efficient stroke width computation (best viewed in color).

Note the color variation of non-characters and characters in (c). Larger color variation indicates larger stroke width variance. (a) Detected regions.(b) Skeleton. (c) Distance transform.

4.2.4 Perceptual Divergence (PD)

A color contrast is a widely adopted measurement of saliency. For the task of scene text detection, there is observed that, in order to ensure reasonable readability of text to a human, the color of text in natural scenes is typically distinct from that of the surrounding area. Thus, the PD cue to measure the perceptual divergence of a region r against its surroundings, which is defined as:

$$PD(r) = \sum_{R,G,B} \sum_{j=1}^{b} h_j(r) \log \frac{h_j(r)}{h_j(r^*)} \quad \ldots(2)$$

Where,

$h_j(r) \log \frac{h_j(r)}{h_j(r^*)}$= measures the dissimilarity of twoprobability distributions in the information

theory

h(r) = color histogram of region

h(r*) = color histogram of region outside r but within its bounding box

$\{j\}_1^b$= index of histogram bins

Note that the more different the two histograms are, the higher the PD is. However, using the intensity channel only ignores valuable color information, which will lead to a reduction in the measured perceptual divergence between distinct colors with the same intensity. In contrast, all three sub-channels (*i.e.*, R, G, B) are utilizedin the computation of perceptual divergence in our approach.

Histogram of Gradients at Edges (eHOG)

The Histogram of Gradients (HOGs) is an effective feature descript or which captures the distribution of gradient magnitude and orientation. There is proposed a characterness cue based on the gradient orientation at edges of a region, denoted by eHOG. This cue aims to exploit the fact that the edge pixels of characters typically appear in pairs with opposing gradient directions.

Firstly, edge pixels of a region *r* are extracted by the Canny edge detector. Then, gradient orientations θ of those pixels are quantized into four types,

Type 1: $0 < \theta \leq \pi/4$ or $7\pi/4 < \theta \leq 2\pi$

Type 2: $\pi/4 < \theta \leq 3\pi/4$

Type 3: $3\pi/4 < \theta \leq 5\pi/4$

Type 4: $5\pi/4 < \theta \leq \pi/4$

Fig.4 .eHOG representation. Sample text (left) and four types of edge points represented in four different colors (right). Note that the number of edge points in blue is roughly equal to that in orange, and so for green and crimson.

An example demonstrating the four types of edge pixels for text is shown in Fig.4. (right), where four different colors are used to depict the four types of edge pixels. As it shows, we can expect that the number of edge pixels in Type 1 should be close to that in Type 3, and so for Type 2 and Type 4.

Based on this observation, we define the eHOG cue as:

…(3)

$$eHOG(r) = \frac{\sqrt{(w_1(r) - w_3(r))^2 + (w_2(r) - w_4(r))^2}}{\sum_{i=1}^{4} w_i(r)}$$

Where,

wi(r) – number of edge pixels in type *i* within region *r*

$\sum_{i=1}^{4} w_i(r)$ – sake of scale invariance

5. APPLICATIONS

5.1. Computer Vision Applications

Detecting text in natural scene images, as opposed to scans of printed pages, faxes and business cards is an important step for a number of computer vision applications such as computerized aid for visually impaired, automatic geo coding of businesses and robotic navigation in urban environments. Retrieving texts in both indoor and outdoor environments provides contextual clues for a wide variety of vision tasks.

5.2. OCR System

It is the mechanical or electronic conversion of images of typewritten or printed text into machine-encoded text. It is widely used as a form of data entry from printed paper data records, whether pass port documents, invoices, bank statements, receipts, business cards, mail or other documents. It is a common method of digitizing printed texts so that it can be electronically edited, searched, stored more compactly, displayed online and used in machine processes such as machine translation, text to speech, key data and text mining.

5.3. Document Image Text Detection

Document image text detection contains to categorize the region of interest in the scanned image of a text document. A reading system requires the segmentation of text zones from no textual ones and the arrangement in their correct reading order. Many document images are rich in color and have complex background. To detect text from them, a standard approach always use both color and binary information.

5.4. Image Retrieval

Image retrieval system is a computer system for browsing, searching and retrieving images from a large database of digital images. Image search is a specialized data search used to find images. To search for images, a user may provide query terms such as keyword.

5.5.Automatic number plate recognition

There are various types of different style of license plate of vehicle. They are used by various police forces and as a method of electronic toll collection on pay per use road and cataloging the movement of traffic or individuals. It has to consider poor resolution, blurry images, poor lighting, low contrast and different font and sizes. On some cars or bikes, two bars may obscure one or two characters of the license plate.

5.6. Text detective in mobile

Text detective is an application for the mobile that can detect text and read it aloud- whether it's a nearby sign, a handout, or any other document. The application uses the phone's camera, and planted text detection algorithm analyze the video stream an tells whether there is text or not.

6. CONCLUSIONS

In this work, there is proposed a scene text detection approach based on measuring 'characterness'. The proposed characterness model reflects the probability of extracted regions belonging to character, which is constructed via fusion of novel characterness cues in the Bayesian framework. In the character labeling model, by constructing a standard graph, not only characterness score of individual regions is considered, similarity between regions is also adopted as the pair wise potential. But there is some problem when applying this method on two types of images. Two kinds of characters that approach cannot handle where characters in extremely blur and low resolution and second one where characters in uncommon fonts. Compared with state-of-the-art scene text detection approaches, we have seen that proposed method is able to achieve accurate and robust results of scene text detection.

REFERENCES

[1] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," IEEE Trans. Pattern Anal. Mach. Intell., vol. 20, no. 11, pp. 1254–1259, Nov. 1998.

[2] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide baseline stereo from maximally stable extremal regions," in Proc. Brit. Mach. Vis. Conf., 2002, pp. 384–393.

[3] L. Neumann and J. Matas, "A method for text localization and recognition in real-world images," in Proc. Asian Conf. Comput. Vis., 2010, pp. 770–783.

[4] B. Alexe, T. Deselaers, and V. Ferrari, "What is an object?" in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2010, pp. 73–80.

[5] B. Epshtein, E. Ofek, and Y. Wexler, "Detecting text in natural scenes with stroke width transform," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2010, pp. 2963–2970.

[6] Manoj Kumar, Young Chul Kim, Guee Sang Lee, "Text Detection using Multilayer Separation in Real Scene Images" in Proc. 10th IEEE Conf.Comput. and Inf. Tech. 2010

[8] C. Yao, X. Bai, W. Liu, Y. Ma, and Z. Tu, "Detecting texts of arbitrary orientations in natural images," in Proc. IEEE Conf. Comput. Vis. Pattern Recognition., Jun. 2012, pp. 1083–1090.

[7] D. Küttel and V. Ferrari, "Figure-ground segmentation by transferring window masks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2012, pp. 558–565.

[8] Y. Li and H. Lu, "Scene text detection via stroke width," in Proc. IEEE Int. Conf. Pattern Recognit., Nov. 2012, pp. 681–684.

[9] Q. Meng, Y. Song, "Text detection in natural scenes with salient region," in Proc. IEEE 10th Int. Workshop Document Anal. Syst.,Mar. 2012, pp. 384–388.

[10] A. Shahab, F. Shafait, A. Dengel, and S. Uchida, "How salient is scene text?" in Proc.