ISSN Print: 2394-7500
ISSN Online: 2394-5869
Impact Factor: 5.2
IJAR 2016; 2(7): 795-798
www.allresearchjournal.com
Received: 23-05-2016
Accepted: 24-06-2016

Somenath Chakraborty
Ph.D. Scholar in the
Department of Computer
Science & Engineering,
University of Calcutta,
Kolkata, India.

Samir K Bandyopadhyay
Professor of the Department of
Computer Science Engineering,
University of Calcutta,
Kolkata, India.

Correspondence
Somenath Chakraborty
Ph.D. Scholar in the
Department of Computer
Science & Engineering,
University of Calcutta,
Kolkata, India.

Scene text detection using modified histogram of oriented gradient approach

Somenath Chakraborty and Samir K Bandyopadhyay

Abstract
With the rapid use of smart phones, digital camera etc we are in a era where it becomes a daily habit to click photos and post it in the internet. The amount of camera capture photos is increasing exponentially and in many photos there is text in it. With the verity of captured images it becomes very difficult to extract text for desired uses. This paper proposes a modified histogram of oriented gradient feature extraction model for detection of text from camera capture images as well as born digital images.

Then we use svm-light for classification of the pattern of the text and our model efficiently performs well on both types of images.

Keywords: Scene text detection, SVM, natural scene text, texture analysis

1. Introduction

Today with the rapid growth of internet, the volume of data is increasing exponentially. There are so many images in the internet which increase the internet data volume and in the form of natural images [11-14]. We are here interested in scene with textual information. Scene text detection is still a challenging task due to factors includes complex background, low quality images, low light conditions and lighting variations, variations of text content and deformation of text appearances, Complex background etc. In this paper we proposed a scheme for generation of features which is robust to face all the modern challenges of today [15-19].

2. Review Research Paper

Scene text detection is of two types. Connected Component analysis and Sliding window based classification.

The component based methods often use colour [1-3], point [19], edge/gradient [20], stroke [21], texture [22], and/or region [23-26] features or a hybrid of them [27-28] to localize text components. The technique used in Connected Component analysis is based on finding homogeneous colour regions present in the text part. Like Maximally Stable External Regions (MSER). Now a days MSER are very successfully use to detect the text region of an image, but as it integrates all image data channels, like Hue (H), Saturation(S), Intensity (I) and Gradient together which increases the complexity and as it basically deals with regular region it lakes crucial understanding in irregular shape which is enormously present in natural scenes.

Sliding window methods use discriminative models to detect text with a multi-scale sliding window classification. Like Scale Invariant Feature Transform (SIFT) by David G. Lowe [5], as it transform images into scale-invariant coordinates relative to local features. SIFT features are first extracted from a set of reference images and stored in a database. The basic technique behind SIFT is scale-space. It is a strong affine invariant feature generation algorithm and its scale invariant property working very usefully in challenging cases too. There are other algorithm is there like PCA-SIFT, GSIFT, CSIFT, SURF and ASIFT which are basically the different variant and improvements of SIFT [5-10].

Another very important feature detector in this class is Histogram of Oriented Gradient (HOG) by Navneet Dalal and Bill Triggs [1] for “pedestrian detection”. This method is based on recursively examining well-normalized local histograms of image gradient orientations in a very dense array.

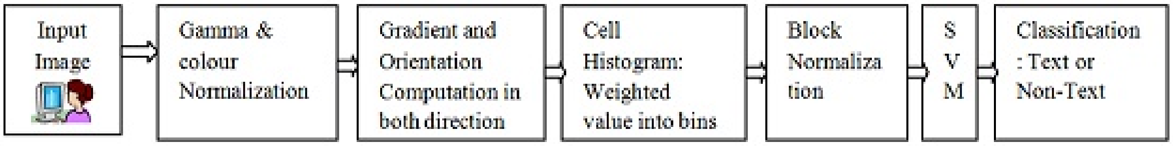
Depending upon there feature calculations, this method is applied in many fields of detection and recognition and even in Scene Text detection as discriminatively trained part-based models by Pedro. F. Felzenszwalb *et al.* [4]. It is part based model where they design the classified which can able to classify the deformable part of the image content.

3. Proposed Approach:

1. In our method of scene text detection we used International Conference on Document Analysis and Recognition (ICDAR) 2011 Robust reading competitions dataset specifically captured separately for Training and Testing.
2. We calculate the modified Histogram of oriented gradient (MHOG) for these ICDAR dataset.
3. Then build the feature values according to the SVM-light multiclass.
4. Evaluate our dataset values both training and testing and obtain results.
5. Depending upon the results we modify the parameter configuration to optimize precision and recall values.

We illustrate the process with an example having image dimension of 40 x 35(Column x Row).

We take the cell size of the image as 2 x 2, So the generated feature vector length is [20 x 18 x 31].This is illustrated in figure 1.


Input Image
Gamma & colour Normalization
Gradient and Orientation Computation in both direction
Cell Histogram: Weighted value into bins
Block Normalization
S V M
Classification : Text or Non-Text

Block Diagram of Text Detection using MHOG Feature Descriptor.

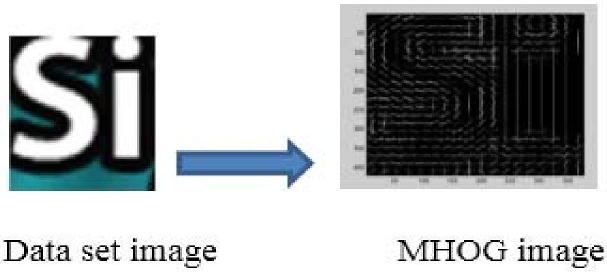


Fig 1: Block of cell be a 2×2 sub-array of cells.

3.1. Block Normalization

In the HOG feature description, Dalal-Triggs [1] describe about four methods for block normalization.

Let v be the non-normalized vector containing all histograms in a given block, $\| v \|_k$ be its k-norm for k=12 and e be some small constant. Then the normalization factor are as follows –

L1- norm: Normalization Factor,

$$f = \frac{v}{||v|| + e}$$

L1-Sqrt: Normalization Factor,

$$f = \sqrt{\frac{v}{||v|| + e}}$$

L2-Norm: Normalization Factor,

$$f = \frac{v}{\sqrt{||v||^2 + e^2}}$$

L2-Hys Norm

L2 –norm followed by clipping and limiting the main value v =0.2. In our case we use L2-norm of block normalization.

For the Dalal-Triggs variant [1], each histogram hd is copied four times, normalised using the four different normalisation factors, the four vectors are stacked, saturated at 0.2, and finally stored as the descriptor of the cell. This results in a num of Orientations * 4 dimensional cell descriptor. Blocks are visited from left to right and top to bottom when forming the final descriptor. They use 13 dimensional features, capturing nine orientations under a single normalization plus four feature capturing without reducing detection accuracy.

In our descriptor (MHOG), we take the gradient orientation of the histogram both in directed manner and undirected manner. This is the main difference from HOG descriptor. As we look into the matter of both undirected and directed orientation, so it is more accurate. By heuristic approach we are taken the window size of feature generation and optimize with the best parameter value which is in our case 35*40 window size where the gradient orientation is very strong to identify the region of interest.

3.2. Classification

We used support vector machine which is basically a supervised learning method for classification. Here we used it to classify two categories of vectors. One is image data which is represented as a vector to feed into the SVM is mainly the portion of text object or foreground object and another one which is also feed to the SVM as the other class, mainly the portion of image which does not contain text in it so it is the back ground object. As SVM gives effective results for classification in high dimensional spaces, here also we obtain the classification of our sample values accordingly.

3.3. Detection Phase

In detection phase the entire image is scan through the scan window detector with the fix, optimized parameter value and the feature is then extracted from the two class of image and feed into the SVM both for the Train image samples and text image samples so that the SVM can train effectively and when we test with an image it can more accurately classify the desired values. It is shown in figure 2.

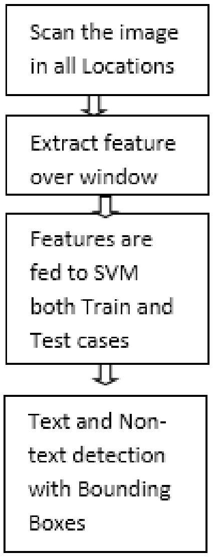

Scan the image in all Locations
Extract feature over window
Features are fed to SVM both Train and Test cases
Text and Non-text detection with Bounding Boxes

Fig 2: Block Diagram of Detection Phase

4. Experimental Results

The Proposed approach was implemented on windows platform, using Open CV and MATLAB 7.6.0(R2008a). Experimental results were obtained on the International Conference on Document Analysis and Recognition (ICDAR) 2011 Robust reading competitions dataset specifically captured separately for Training and Testing.

For the "Train file" we take 840 positive images i.e. image sample that have text in it and 810 negative images i.e. image sample that do not have text in it. So total image sample size for the train category is 1650.For the "Test file" we take 204 positive sample images and 204 negative sample images. So total image sample size in this category is 408. After getting results from the linear SVM we recursively optimize the regularization parameter and kernel values so that we achieve the desired values in a heuristic manner. Depending upon the values we prepare the confusion matrix by which Precision and recall is calculated. We obtained the two different degree known as precision or Positive predictive value (p) and recall or sensitivity(r) of our experimental results. Expressions used for computation of p and r values are as follows –

Precision: Percentages of image data values that the classified labelled as positive are actually positive.

So,

$$\text{Precision} = \frac{TP}{TP+FP}$$

TP: True positive, FP: False positive

Recall: Percentages of image data values the classified labelled as positive.

So,

$$\text{Recall} = \frac{TP}{TP+FN}$$

TP: True positive, FN: False Negative

The results are presented with other famous methods in Table 1 along with text detection results on a few images as shown in figure 3.

Table 1: Comparative Performance analysis of Text Detection Approaches.

Algorithms	Precision(p)	Recall(r)
Roy Chowdhuri *et al.*[28]	0.57	0.59
Kasar *et al.*[29]	0.63	0.59
Epshtein *et al* [30]	0.73	0.60
Chen *et al* [24]	0.73	0.60
Merino-Gracia *et al.* [31]	0.51	0.67
Zhang and Kasturi *et al.* [32]	0.67	0.46
Proposed method	0.75	0.68

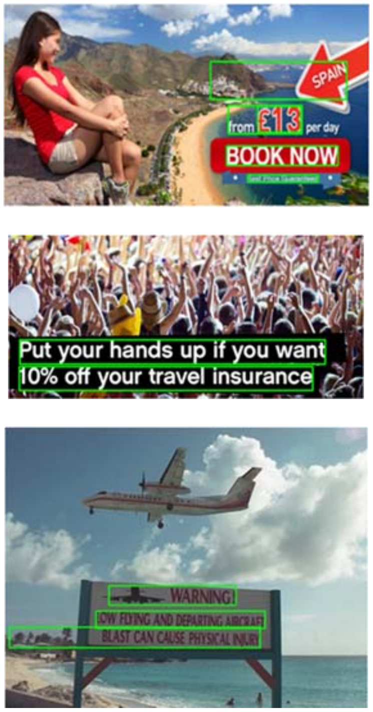

SPAIN
from £13 per day
BOOK NOW
Put your hands up if you want
10% off your travel insurance
WARNING
LOW FLYING AND DEPARTING AIRCRAFT
BLAST CAN CAUSE PHYSICAL INJURY

Fig 3: Text detection from images

5. Conclusion and Future work

This paper proposed the method (MHOG) as robust, affine method for generation of object features for detection of text in scene images mainly captured by digital camera and born digital images but it can be applied in many field like licence plate detection of car, category classification etc. We have a plan to use this in all possible areas of applications.

6. References

1. Dalal N, Triggs B. Histograms of Oriented Gradients for Human Detection. In Proc. CVPR, 2005.
2. Crammer K, Singer Y. On the Algorithmic Implementation of Multi-class SVMs, JMLR, 2001.
3. Yi C, Tian Y. Localizing Text in Scene Images by Boundary Clustering, Stroke Segmentation and String Fragment Classification, IEEE Trans. Image Processing, 2012; 21(9):4256-4268.
4. Felzenszwalb PF, Girshick RB, McAllester D, Ramanan D. Object detection with discriminatively trained part-based models, IEEE Trans. Pattern Anal. Mach. Intell., 2010; 32(9):1627-1645.

5. Lowe DG. Distinctive image features from scale-invariant key points. International Journal of Computer Vision. 2004; 60(2):91-110,
6. Viola P, Jones MJ. Robust real-time face detection. International Journal of Computer Vision. 2004; 57(2):137-154, ISSN 0920-5691.
7. Chen X, Yuille AL. Detecting and reading text in natural scenes. CVPR, 2004; 2:366-377,
8. Bishop C. Pattern recognition and machine learning. Springer, 2006.
9. Jain AK, Yu B. Automatic text location in images and video frames, Proc. ICPR, 1998, 1497-1499. N Sharma, U Pal, Blumenstein M. Recent advances in video based document processing: A review, in: Proc. IAPR International Workshop on Document Analysis Systems, IEEE Computer Society, Los Alamitos, CA, USA, 2012, 63-68.
10. Yi C, Tian Y. Text string detection from natural scenes by structure-based partition and grouping, IEEE Trans. on Image Processing. 2011; 20(9):2594-2605.
11. Shivakumara P, Huang W, Phan TQ, Tan CL. Accurate video text detection through classification of low and high contrast images, Pattern Recognition, Elsevier. 2010; 43(6):2165-2185.
12. Joachims T. Making large-scale svm learning practical, in Advances in Kernel Methods - Support Vector Learning, B. Sch¨olkopf, C. Burges, and A. Smola, Eds. MIT Press, 1999.
13. Ke Y, Sukthankar R. PCA-SIFT: A more distinctive representation for local image descriptors, in IEEE Conference on Computer Vision and Pattern Recognition, 2004.
14. Winder SA, Brown M. Learning Local Image Descriptors, in Computer Vision and Pattern Recognition, CVPR ’07. IEEE Conference on, 2007, 1-8.
15. Neumann L, Matas J. A method for text localization and recognition in real-world images, Computer Vision-ACCV 2010, 2011, 770-783.
16. Mikolajczyk K, Schmid C. A Performance Evaluation of Local Descriptors, IEEE Trans. Pattern Analysis and Machine Intelligence. 2005; 27(10):1615-1630.
17. Vedaldi A, Fulkerson B. VLFeat: An open and portable library.http://www.vlfeat.org/, 2008.
18. Grauman K, Darrell T. The Pyramid Match Kernel: Efficient Learning with Sets of Features, J Machine Learning Research. 2007; 8:725-760.
19. Zhao X, Lin KH, Fu Y, Hu Y, Liu Y, Huang TS. Text from Corners: A Novel Approach to Detect Text and Caption in Videos, IEEE Trans. Image Processing. 2011; 20(3):790-799.
20. Phan TQ, Shivakumara P, Tan CL. Text Detection in Natural Scenes Using Gradient Vector Flow-Guided Symmetry, Proc. IEEE Int’l Conf. Pattern Recognition, 2012, 3296-3299,
21. Epshtein B, Ofek E, Wexler Y. Detecting Text in Natural Scenes With Stroke Width Transform, Proc. IEEE Int’l Conf. CVPR, 2010.
22. Ye Q, Huang Q, Gao W, Zhao D. Fast And Robust Text Detection in Images and Video Frames, Image and Vision Computing. 2005; 23:565-576.
23. Neumann L, Matas J. Text Localization in Real-world Images using Efficiently Pruned Exhaustive Search, Proc. Int’s Conf. on Document Analysis and Recognition, 2011, 687-691.
24. Chen H, Tsai SS, Schroth G, Chen DM, Grzeszczuk R, Girod B. Robust Text Detection in Natural Images with Edge-Enhanced Maximally Stable Extremal Regions, Proc. IEEE Int’l Conf. Image Processing, 2011, 2609-2612.
25. Neumann L, Matas J. Real-time Scene Text Location and Recognition, Proc. IEEE Int’l Conf. CVPR, 2012, 3538-3545,
26. Koo DH, Kim, Scene Text Detection via Connected Component Clustering and Non-text Filtering, IEEE Trans. Image Processing. 2013; 22(6):2296-2305.
27. Yi C, Tian Y. Text String Detection from Natural Scenes by Structure-Based Partition and Grouping, IEEE Trans. Image Processing. 2011; 20(9):2594-2605,
28. Chowdhury AR, Bhattacharya U, Parui SK. Text detection of two major Indian scripts in natural scene images, Proc. CBDAR, Lecture Notes in Comp. Sci. 2012; 7139:42-57,
29. Kasar T. Font and background color independent text binarization, Proc. of CBDAR, 2007, 3-9.
30. Epshtein B, Ofek E, Wexler Y. Detecting Text in Natural Scenes with Stroke Width Transform. Proc. of CVPR, 2010, 2963-2970.