21st International Conference on Pattern Recognition (ICPR 2012)
November 11-15, 2012. Tsukuba, Japan

Text Location in Complex Images

Álvaro González, Luis M. Bergasa, J. Javier Yebes, Sebastián Bronte
Dept. of Electronics, University of Alcalá, Spain
{*alvaro.g.arroyo, bergasa, javier.yebes, sebastian.bronte*}*@depeca.uah.es*

Abstract

An automatic text recognizer needs, in first place, to localize the text in the image the more accurately possible. For this purpose, we present in this paper a robust method for text detection. It is composed of three main stages: a segmentation stage to find character candidates, a connected component analysis based on fast-to-compute but robust features to accept characters and discard non-text objects, and finally a text line classifier based on gradient features and support vector machines. Experimental results obtained with several challenging datasets show the good performance of the proposed method, which has been demonstrated to be more robust than using multi-scale computation or sliding windows.

1. Introduction

Automatic text recognition is one of the hardest problems in computer vision. An essential prerequisite for text recognition is to robustly locate the text on the images. Nevertheless, this still remains a challenging task because of the wide variety of text appearance due to variations in font, thickness, color, size, texture, and also geometric distortions, partial occlusions, different lighting conditions and image resolutions.

In order to assess the state of the art in text location, the Robust Reading Competition has been recently held in the frame of the ICDAR 2011 conference. Two challenging public datasets were released for this competition. In this work, we evaluate the performance of our proposed system with both datasets. The results show that our proposed method is really competitive.

The main contributions of this paper are, in first place, a new segmentation method based on a combination of MSER and a locally adaptive thresholding method, and secondly, a thorough study on different simple and fast-to-compute features to distinct text from non-text. Section 2 describes this study, while section 3 briefly explains the text location algorithm. Section 4 provides the results and section 5 concludes the paper.

2. Text features analysis

In order to obtain a set of distinctive features capable of distinguishing character objects from non-character objects, we have made an analysis of certain text features under the ICDAR 2003 train dataset. Among all the features that we have computed, we find that the more distinctive are those shown in (1)-(8).

$$Occupy\ rate = \frac{area}{height * width} \tag{1}$$

$$Aspect\ ratio = \frac{max(width, height)}{min(width, height)} \tag{2}$$

$$Compactness = \frac{area}{perimeter * perimeter} \tag{3}$$

$$Solidity = \frac{area}{convex\ area} \tag{4}$$

$$Occupy\ rate\ convex\ area = \frac{convex\ area}{height * width} \tag{5}$$

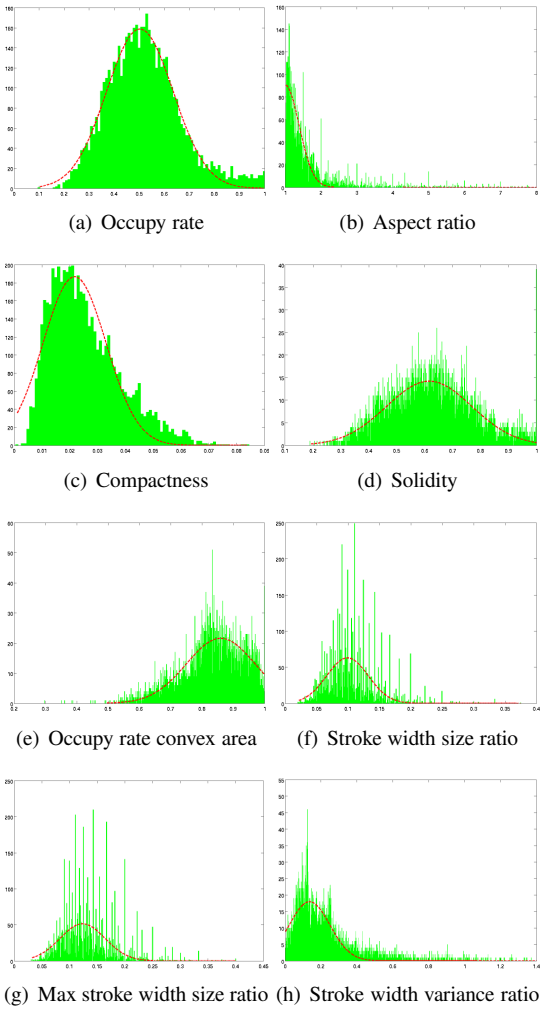
$$Stroke\ width\ size\ ratio = \frac{Stroke\ width}{max(height, width)} \tag{6}$$

$$Max\ stroke\ width\ size\ ratio = \frac{Max\ stroke\ width}{max(height, width)} \tag{7}$$

$$Stroke\ width\ variance\ ratio = \frac{Stroke\ width\ variance}{Stroke\ width} \tag{8}$$

The convex area is the area of the convex hull, which is the smallest convex polygon that contains the region. A stroke is a contiguous part of an image that forms a band of a nearly constant width. Characters are made of strokes which have consistent stroke width. The Stroke Width Transform (SWT) [4] is a local image operator that computes per pixel the width of the most likely stroke containing the pixel.

Fig. 1 and Fig. 2 show the histograms of each of the features in (1)-(8). We see that they follow a Gaussian distribution, or half a Gaussian distribution in case of the aspect ratio, for character components, but it cannot be made the same approximation for non-character components.



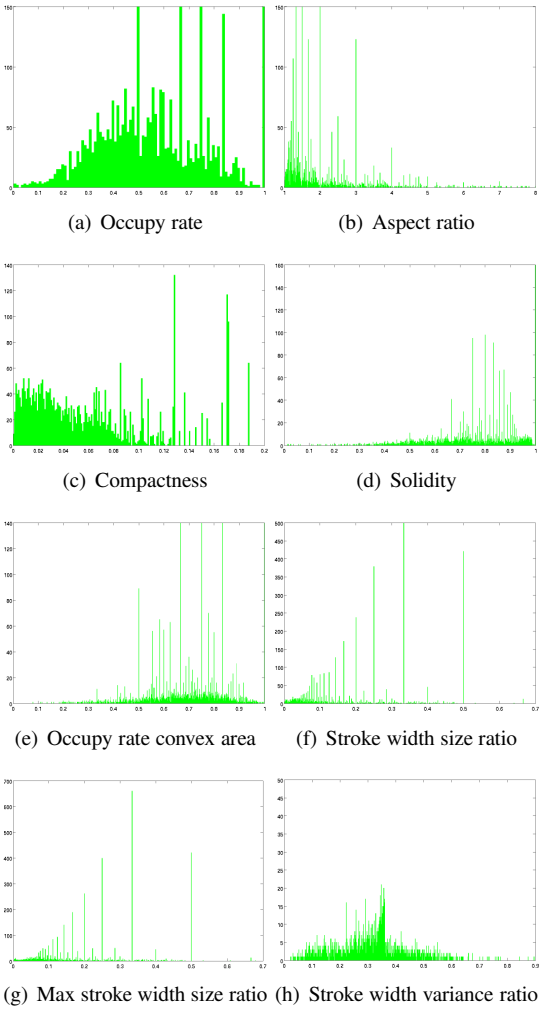
(a) Occupy rate (b) Aspect ratio

(c) Compactness (d) Solidity

(e) Occupy rate convex area (f) Stroke width size ratio

(g) Max stroke width size ratio (h) Stroke width variance ratio

Figure 1. Histograms of features vs approximated Gaussian functions for character components on ICDAR'03 train set.



(a) Occupy rate (b) Aspect ratio

(c) Compactness (d) Solidity

(e) Occupy rate convex area (f) Stroke width size ratio

(g) Max stroke width size ratio (h) Stroke width variance ratio

Figure 2. Histograms of features for non-character components on ICDAR'03 train set.

We have also carried out the same analysis for each character separatedly and we have seen that the histograms of the features can be also approximated by a Gaussian function. The values of the standard deviation of the features for each individual character are, in general, lower than the values obtained for the general case, but not as low as it could be expected. It means that the variability for a single character is almost as large as the variability for all characters altogether.

3. Text location algorithm

The flowchart of our text location algorithm is shown in Fig. 3. Initially, letter candidates are found using a segmentation method that combines the complementary properties of MSER [11] and a locally adaptive thresholding method [15]. Both dark-on-bright and bright-on-dark candidates are extracted in this stage. Then, the resulting candidates are filtered using certain constraints based on the study we have shown in section 2. We reject those objects for which at least one of the features (1)-(8) is out of the range $(\mu_i - 2\cdot\sigma_i, \mu_i + 2\cdot\sigma_i)$, experimentally seen as the optimum one, being μ_i and σ_i the mean and the standard deviation for each feature respectively. The maximum number of holes and the minimum font height allowed are 2 holes and 10 pixels, respectively. Some text candidates can be erroneously rejected, especially those letters which have a high aspect ratio. In order to bring back the mistakenly removed characters, we apply a method to restore them. This method takes into account that adjacent characters are

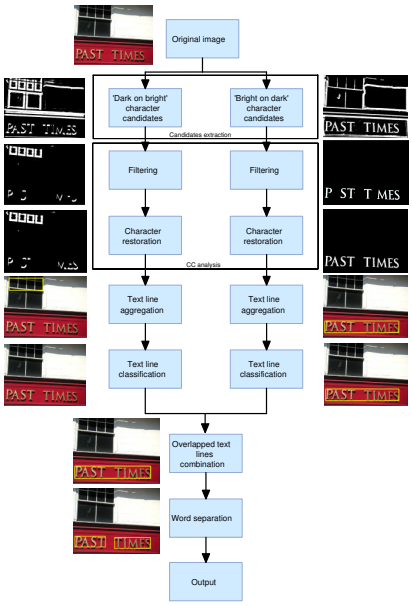


Figure 3. The flowchart of the algorithm

expected to have similar attributes, such as height and stroke width, as well as the Ashida's conditions [10]. Then, character candidates are grouped into lines and each line is classified into text or non-text in order to reject false positives. For this purpose, we use a classifier based on SVM with linear kernel and three different types of features: Mean Difference Feature (MDF) [6], Standard Deviation (SD) and HOG [3]. Finally, words within a text line are separated, giving segmented word areas at the output of the system.

4. Experimental results

We evaluate the proposed method by running it on several public test datasets and comparing to the state of the art. In the following subsections, we show the results obtained for each dataset.

4.1. ICDAR 2003 test dataset

The ICDAR 2003 test dataset has been used as a benchmark for most researchers in the field of text detection in the last decade. Table 1 shows the comparison of our algorithm with the winners of the Robust Reading competitions in ICDAR 2003 and 2005, as well as with some of the methods that have worked with this dataset in the last years. It can be seen that we score second in the global ranking, although we outperform the results obtained in the framework of ICDAR 2003 and 2005 competitions, whose winner was Hinnerk Becker's method.

Table 1. Text localization ICDAR'03 dataset.

Algorithm	Precision	Recall	f
Pan et al. [12]	0.67	0.70	0.69
Our system	**0.81**	**0.57**	**0.67**
Ephstein [4]	0.73	0.60	0.66
H. Chen [2]	0.73	0.60	0.66
Lee et al. [8]	0.69	0.60	0.64
1st ICDAR'05 [9]	0.62	0.67	0.62
Yao [15]	0.64	0.60	0.61
Alex Chen [9]	0.60	0.60	0.58
Zhang & Kasturi [16]	0.67	0.46	0.55
1st ICDAR'03 [10]	0.55	0.46	0.50

4.2. ICDAR 2011 test datasets

In order to assess the state of the art in text location, a new Robust Reading Competition has been recently held in the frame of the ICDAR 2011 conference. Two challenging public datasets were released for this competition, one aimed at reading text in born-digital images [7] and the other one aimed at reading text in scene images [13]. Table 2 and Table 3 show the comparison of our proposed method with the participants in each competition, respectively. We have used the resources available for the competitors in both challenges to compute the performance of our method, *i.e.* the Challenge Web Site for Challenge 1 and the DetEval software [14] for Challenge 2. It can be seen that our method scores first in Challenge 1 and second in Challenge 2.

4.3. CoverDB test dataset

Finally, we have also tested our method with a recent benchmark that contains hundreds of images of CD/DVD cover images [5]. Table 4 shows that we outperform the other methods that have been tested on this dataset.

Table 2. Text localization ICDAR'11 Chall. 1 (%).

Algorithm	Precision	Recall	H. Mean
Our system	**89.23**	**70.08**	**78.51**
Textorter	85.83	69.62	76.88
TH-TextLoc	80.51	73.08	76.62
TDM IACAS	84.64	69.16	76.12
OTCYMIST	64.05	75.91	69.48
SASA	67.82	65.62	66.70
Text Hunter	75.52	57.76	65.46

Table 3. Text localization ICDAR'11 Chall. 2 (%).

Algorithm	Precision	Recall	H. Mean
Kim's method	82.98	62.47	71.28
Our system	**72.67**	**56.00**	**63.25**
Yi's method	67.22	58.09	62.32
TH-TextLoc	66.97	57.68	61.98
Neumann's method	68.93	52.54	59.63
TDM IACS	63.52	53.52	58.09
LIP6-Retin	62.97	50.07	55.78
KAIST AIPR System	59.67	44.57	51.03
ECNU-CCG method	35.01	38.32	36.59
Text Hunter	50.05	25.96	34.19

Table 4. Text localization CoverDB.

Algorithm	Performance
Our system	**0.45**
Escalera et al. [5]	0.28
Cano and Perez [1]	0.16

5. Conclusions

A new method to locate text in images with complex background has been presented. It combines efficiently MSER and a locally adaptive thresholding method. The result is a connected-component-based approach that extracts basic letter candidates using a series of easy and fast-to-compute features. These features, after having been extracted from a challenging train dataset which contains different texts in a huge variety of situations, have proved to follow a Gaussian distribution. It means that they can be used with any dataset independently from their size, color or font. Actually, the proposed method has been tested on four different test datasets and the achieved results show the competitiveness of the method. Unlike other methods, a strong point is the use of feedback in order to restore those characters that might have been filtered out erroneously after computing the text features for each letter candidate. It has been also proposed to use a classifier based on simple features such as mean, standard deviation and HOG computed over image blocks in order to remove repeating structures that can be easily confused to text lines, such as bricks or fences.

Acknowledgments

Work funded through the projects ADD-Gaze (TRA2011-29001-C04-01), Ministerio de Economía y Competitividad, and Robocity2030 (CAM-S-0505/DPI000176), Comunidad de Madrid.

References

[1] J. Cano and J. C. Perez-Cortes. Vehicle license plate segmentation in natural images. *Lecture Notes on Computer Science 2652*, pages 142–149, 2003.

[2] H. Chen, S. Tsai, G. Schroth, D. Chen, R. Grzeszczuk, and B. Girod. Robust text detection in natural images with edge-enhanced maximally stable extremal regions. In *ICIP*, 2011.

[3] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *CVPR*, 2005.

[4] B. Ephstein, E. Ofek, and Y. Wexler. Detecting text in natural scenes with stroke width transform. In *CVPR*, 2010.

[5] S. Escalera, X. Baro, J. Vitria, and P. Radeva. Text detection in urban scenes. *Frontiers in Artificial Intelligence and Applications*, 202:35–44, 2009.

[6] S. M. Hanif and L. Prevost. Text detection and localization in complex scene images using constrained adaboost algorithm. In *ICDAR*, 2009.

[7] D. Karatzas, S. R. Mestre, J. Mas, F. Nourbakhsh, and P. P. Roy. ICDAR 2011 Robust Reading Competition. Challenge 1: Reading Text in Born-Digital Images (Web and Email). In *ICDAR*, 2011.

[8] S. H. Lee, M. S. Cho, K. Jung, and J. H. Kim. Scene text extraction with edge constraint and text collinearity. In *ICPR*, 2010.

[9] S. Lucas. Text locating competition results. In *ICDAR*, 2005.

[10] S. M. Lucas, A. Panaretos, L. Sosa, A. Tang, S. Wong, R. Young, K. Ashida, H. Nagai, M. Okamoto, H. Yamamoto, H. Miyao, J. Zhu, W. Ou, C. Wolf, J.-M. Jolion, L. Todoran, M. Worring, and X. Lin. ICDAR 2003 robust reading competitions: entries, results, and future directions. *IJDAR*, 7(2-3):105–122, 2005.

[11] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide baseline stereo from maximally stable extremal regions. In *BMVC*, 2002.

[12] Y. F. Pan, X. Hou, and C. Liu. A hybrid approach to detect and localize texts in natural scene images. *IEEE Trans. on Image Processing*, 20(3):800–813, 2011.

[13] A. Shahab, F. Shafait, and A. Dengel. ICDAR 2011 Robust Reading Competition. Challenge 2: Reading Text in Scene Images. In *ICDAR*, 2011.

[14] C. Wolf and J.-M. Jolion. Object count/area graphs for the evaluation of object detection and segmentation algorithms. *Int. Jrnl. on Document Analysis and Recognition*, 8(4):280–296, 2006.

[15] J. L. Yao, Y. Q. Wang, L. B. Weng, and Y. P. Yang. Locating text based on connected component and SVM. In *ICWAPR*, 2007.

[16] J. Zhang and R. Kasturi. Text detection using edge gradient and graph spectrum. In *ICPR*, 2010.