

Classification of regions extracted from scene images by morphological filters in text or non-text using decision tree

Wonder Alexandre Luz Alves
Department of Computer Science
Institute of Mathematics and Statistics
University of São Paulo, Brazil
wonder@ime.usp.br

Ronaldo Fumio Hashimoto
Department of Computer Science
Institute of Mathematics and Statistics
University of São Paulo, Brazil
ronaldo@ime.usp.br

Abstract

We present in this work a new method to classify regions extracted from scene images by morphological filters in text or non-text region using a decision tree. Our technique can be divided into three parts. Firstly, we extract a set of regions by a robust scheme based on morphological filters. Then, after a refinement, a set of text attributes is obtained for each region. In the last step, a decision tree is built in order to classify them as text or non-text regions. Experiments performed using images from the ICDAR public dataset show that this method is a good alternative for practical problems involving text location in scene images.

Keywords: text region classification, text localization, text information extraction, morphological filters.

1 INTRODUCTION

In the last years, several algorithms for *text information extraction* (TIE) from scene images have been proposed [4, 9, 7, 12, 19, 18]. In spite of such studies, it is not easy to design a general-purpose TIE system [11] for scene images, since texts that are present in these type of images are considered as an integral part of the scene and their presence are almost always accidental and non-intentional. Owing to this factor, text occurrences in these scene images can be significantly different from one another with respect to slopes, sizes, font styles, illumination and also they can be partially occluded. In Fig. 1, we show some examples of scene images.



Figure 1: Scene Images extracted from the ICDAR dataset [13].

According to Jung et al. [11], the TIE problem can be divided into five subproblems: (1) detection; (2) localization; (3) tracking; (4) extraction and enhancement; and (5) recognition. Our work focuses on the second subproblem whose main objective is to locate text regions within an input image. Since text regions usually are composed by characters aligned along a line such that for any two adjacent characters there is a high contrast with a uniform background, we can consider, as the first step for solving the problem of text region localization, the problem of localizing region candidates that contain uniform background and similar shapes with high contrast with respect to their background that are aligned along a line. As the second step, we can label these candidates into text or non-text regions using a built classifier. Thus, let us define a *text region candidate* in the following way.

Definition 1 (Text Region Candidate) Let ϵ be a small positive real number. A text region candidate R is defined as the smallest involving rectangle with a uniform background that contains objects with similar shapes and high contrast (with respect to their background) positioned along a line such that for any two adjacent objects C_i and C_{i+1} belonging to R , we have that $\text{dist}(C_i, C_{i+1}) < \epsilon$, where dist is a distance function.

Text location methods can be classified into two types: texture-based [4, 12] and region-based [6, 9, 7, 18, 19]. The texture-based methods assume that the character attributes are different from their background in the text region making them possible to be discriminated. Techniques such as Gabor filter, wavelets, fast Fourier transform can be used to extract text attributes from potential text

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Copyright UNION Agency – Science Press

regions in order to classify them. The region-based methods [6, 9, 7, 18, 19] use several attributes already present in the image. In general, their procedure can be divided into three parts: (i) obtain a set of region candidates by emphasizing the high contrast between text and their background; (ii) then, merge the obtained candidates by a geometrical analysis of their spatial arrangements; (iii) and finally, determine if the obtained regions are text or non-text using heuristics or classifiers.

In this work, we present a new region-based method to classify regions extracted from scene images by morphological filters in text or non-text region using a decision tree. Our technique can be divided into three parts. Firstly, a set of region candidates (connected components) is selected by using a robust scheme based on morphological filters. A refinement of the obtained region candidates is performed in order to emphasize regions with text presence. Then, a set of text attributes is obtained for each region. Finally, a decision tree is built in order to classify them as text or non-text regions. Experiments performed using images from the ICDAR public dataset show that this method is a good alternative for practical problems involving text location in scene images.

After this short introduction, Section 2 presents a brief review of some related works with the commonly main hypotheses used for text region localization. Then, our method is presented in Section 3. In Section 4, we show the experimental results. Finally, Section 5 concludes this work and points out some future research.

2 RELATED WORK

Region-based methods usually have the strategy of selecting a set of region candidates and then eliminate the non-text regions. In this context, morphological filters have been widely used to extract set of text region candidates [6, 7, 9, 18, 19]. Once these candidates are extracted, some methods use heuristics [6, 9, 19], others classifiers [10, 18] to select the text regions.

Wu et al. [19] proposed a scheme based on morphological filters to locate candidates for text regions based on the high contrast between texts and their background, and afterwards they applied some heuristics to find the actual text regions. Hasan and Karam's method [9] uses morphological filters to extract region candidates and then non-text regions are eliminated by heuristics. Lixu et al. [7, 6] proposed a method based on the difference of top-hat transforms to segment region candidates and then subsequently text regions are reconstructed using conditional dilations. Retornaz and Marcotegui [18] developed a method based on ultimate opening to extract the characters from text regions, and thereafter they use a classifier based on linear Fisher discriminant to classify text regions. Renjie et al. [10]

proposed a two stage text region classification: (i) eliminate impossible regions using heuristics; (ii) and then separate text from non-text regions using an SVM classifier. The same ideas to classify text regions are used in [3, 8, 18].

This work, as well as others [9, 10, 18, 19], assumes as true some hypotheses that take account the contrast between text regions and their background and the geometric regularity of the font within the text region. Such hypotheses (called here as *text region hypotheses*) are classified as following:

1. Contrast

- (a) There is a contrast between text region and its background.
- (b) The text gray levels within the same region are similar.

2. Font Geometry

- (a) Characters within the same region:
 - they have similar heights and widths.
 - they are aligned along a line.
 - distances between any two adjacent characters are similar.

3. Prior Knowledge

- (a) A text region consists of at least three characters.

3 PROPOSED METHOD

The proposed method consists of three stages (see Fig. 2): (i) extraction of regions; (ii) selection of regions; and (iii) classification of the selected regions in text and non-text.

Extraction of regions: In this stage, a set of region candidates is extracted from the input image using morphological filters.

Selection of regions: The next stage consists of refining the extracted regions in order to emphasize the most probable text regions. This is done by using heuristics based on the text region hypotheses.

Classification of regions: The last stage of the method consists of obtaining a set of features from the selected regions that will be later used as an input to a decision tree in order to classify them as text or non-text regions.

3.1 Extraction of Region Candidates

The extraction of region candidates from the input image must be orientation invariant and also highly noise tolerant. It is well know that morphological operators can be successfully used to accomplish these tasks. In Fig. 3, we show a simple flow chart of our extraction procedure.

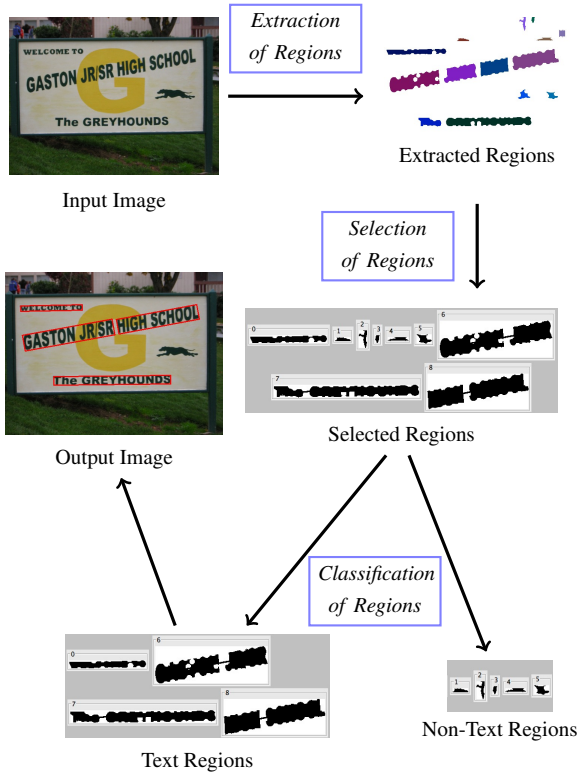


Figure 2: Overview of the proposed method.

The extraction procedure of the region candidates (based on the contrast hypotheses) can be briefly described by the following steps. (1) Firstly, if the input f is a color image, then it is converted into a gray scale image using the following equation [5]:

$$f(x) = [0.299 \cdot f_r(x) + 0.587 \cdot f_g(x) + 0.114 \cdot f_b(x)],$$

where f_r, f_b, f_g are the three RGB components of the input image. (2) Then, an opening and a closing top-hats are applied to the image f using as the structuring element (SE) the disk of radius λ , producing two output images f_w and f_c , respectively. (3) Afterwards, build the image f_m by taking the maximum between f_w and f_c pixel-by-pixel. The image f_m contains all the characters with thickness less than or equal to the radius λ of the SE. (4) The next step is to apply the closing to the f_m by a 3×3 square SE in order to close small borders within the extracted regions. (5) In the sequence, the image f_m is binarized by the local Otsu's thresholding method [15] obtaining the binary image f_a . (6) Then, an area-opening filter is applied to f_a (producing the binary image f_b) to eliminate connected components smaller than the area of the disk of radius λ . (7) Thereafter, the image f_b is decomposed into n connected components $\Lambda(f_b) = \{R_1, R_2, \dots, R_n\}$ to be latter used in the next parts. In Fig. 4, we present an example of the application of extraction procedure.

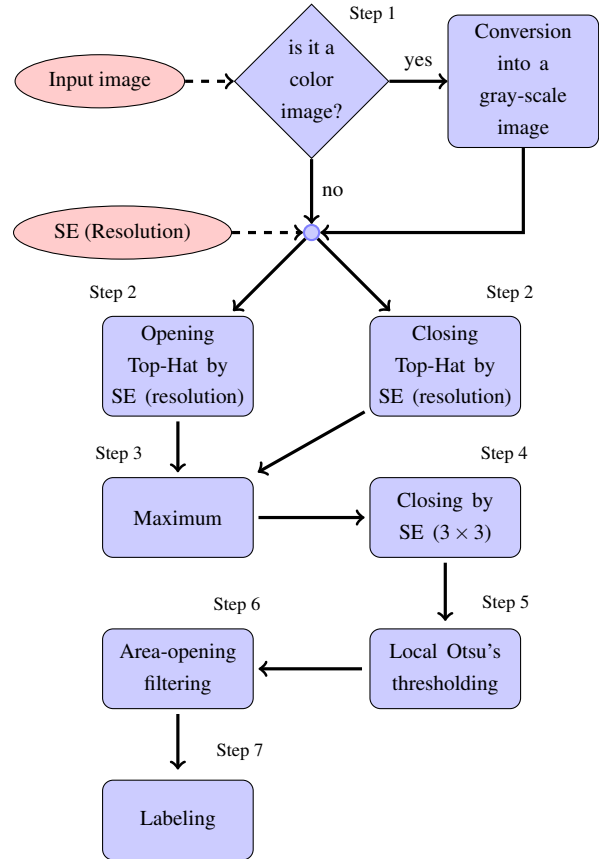


Figure 3: A simple flow chart showing our procedure for extraction of candidate regions.

3.2 Selection and Merging of Regions

After obtaining the set $\Lambda(f_b)$ of region candidates from the input image f , a subset $\zeta \subseteq \Lambda(f_b)$ with the potential text regions are *selected* based on the contrast and geometric hypotheses. Afterwards, a geometric analysis is performed on the selected regions to *merge* the ones that may belong to the same text region.

Selection of Regions Based on the rectangularity criteria [9, 19], we build a subset $\zeta \subseteq \Lambda(f_b)$ containing the potential text regions. Let $R \in \Lambda(f_b)$ be a connected component candidate for a text region and let R^θ denote the rotated version of R along its longest axis [5]. The connected component R is said to be a text region candidate if the following statements hold:

1. the density must be between 0.2 and 0.95, that is,

$$0.2 < \frac{|R|}{W_{R^\theta} \cdot H_{R^\theta}} < 0.95;$$

2. the width of the rotated connected component R^θ must be larger than its height, that is,

$$\frac{W_{R^\theta}}{H_{R^\theta}} > 1.5,$$

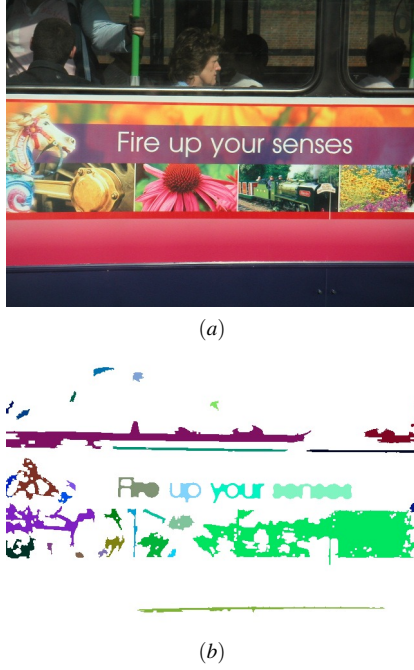


Figure 4: Extraction of region candidates using morphological filters. (a) Input image; (b) connected components obtained by the application of the extraction region candidate procedure.

where $|R|$ denotes the area of the component R ; W_{R^θ} and H_{R^θ} are the width and height of the rotated connected component R^θ along the direction θ , respectively. The first criterion eliminates small and large components that are impossible to be text regions; while the second criterion is a characteristic that must be satisfied by the most characters. These two criteria (based on the works [9, 19]) can eliminate a large number of regions that could not be text regions.

Merging of Regions Two distinct components $R_i, R_j \in \Lambda(f_b)$ belong to the same text region if their orientations, heights and alignments are similar. So, R_i and R_j are in the same text region if the following statements hold:

1. The difference between their orientations must be less than 15 degrees, that is,

$$d_\theta(R_i, R_j) = \min \left\{ \begin{aligned} &|\theta_{R_i} - \theta_{R_j}|, \\ &|\theta_{R_i} - \theta_{R_j} + 360|, \\ &|\theta_{R_i} - \theta_{R_j} - 360| \end{aligned} \right\} < 15, \quad (1)$$

where θ_{R_i} and θ_{R_j} are the orientations of R_i and R_j , respectively. This criterion comes from the work [19].

2. The heights of R_i and R_j must be similar, that is,

$$d_H(R_i, R_j) = |H_{R_i^\theta} - H_{R_j^\theta}| < \min \left\{ H_{R_i^\theta}, H_{R_j^\theta} \right\}, \quad (2)$$

where H_{R^θ} is the height of the rotated component R^θ . This criterion was inspired from the work [18].

3. The third criterion requires that the centroids of R_i and R_j must be similar, that is,

$$d_C(R_i, R_j) = \|CenR_i - CenR_j\|^2 < \max \left\{ W_{R_i^\theta}, W_{R_j^\theta} \right\}, \quad (3)$$

where $CenR$ is the position of the centroid of the component R and W_{R^θ} is the width of the rotated component R^θ . This criterion was inspired from the work [19].

4. Let L_R be the longest axis of R obtained by the equation $y = m_R \cdot x + b_R$. Then, the distance between the axes L_{R_i} and L_{R_j} of R_i and R_j , respectively, can be defined as:

$$d_L(R_i, R_j) = \frac{|y_{CenR_j} - m_{R_i} \cdot x_{CenR_j} - b_{R_i}|}{2\sqrt{1 + m_{R_i}^2}} + \frac{|y_{CenR_i} - m_{R_j} \cdot x_{CenR_i} - b_{R_j}|}{2\sqrt{1 + m_{R_j}^2}},$$

where x_{CenR} and y_{CenR} are the coordinates of the centroid $CenR$. This criterion requires that

$$d_L(R_i, R_j) < \min \left\{ H_{R_i^\theta}, H_{R_j^\theta} \right\}. \quad (4)$$

This criterion was inspired from the work [19].

Although all these four criteria come from other works [18, 19] (and consequently, they are not so original), the thresholds given by the right side of the last three inequalities (Eqs. 2, 3 and 4) are computed using the values taken from regions instead of single characters or constant numbers (differently from the corresponding original works [18, 19]). This approach makes our method more robust when applied to a set of distinct images.

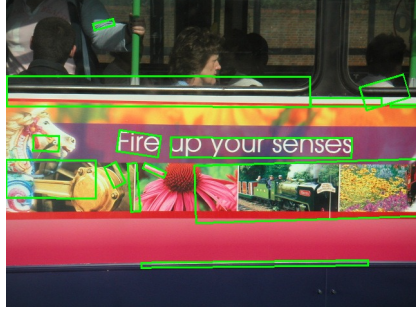
Based on these four criteria, it is possible to decide if R_i and R_j are merged or not, that is, $R' \leftarrow \{R_i, R_j\}$ will be put into ζ and, at the same time, R_i and R_j are taken out from ζ .

The main advantage of applying the merging of regions is that the attributes (features) taken from regions with larger number of characters are likely to give more information to the classifier that will be used later to discriminate text and non-text regions. Besides, text regions that have been fragmented in the extraction of regions stage can be recovered in this stage.

In Fig. 5, we present an example of the application of these criteria applied to the connected components of the image shown in Fig. 4(a).



(a)



(b)

Figure 5: Results by the application of the selection and merging region candidates. (a) Image containing the labeled connected components of $\Lambda(f_b)$ after candidates extraction; (b) image showing (marked by green rectangles) the selected regions obtained from ζ .

3.3 Classification of Text Region by Decision Tree

After selecting a set ζ that contains all the potential text regions, the next step consists of building a classifier to discriminate whether a given element $R \in \zeta$ is a text or a non-text region. In the following, we will describe how to obtain features from the connected components $R \in \zeta$ that will be later used for classification. Most ideas presented in this section for feature extraction have been taken from the work [9, 19]. Differently from [9, 19], we use these features to construct a vector that will be later used as the input for the decision tree classifier.

Features Extraction The features considered for text and non-text region classification are extracted based on the hypotheses described in Section 2. Given a region $R \in \zeta$, the x -projection can be used to extract geometric features from R [1, 19]. If R is an actual text region, then the characters presented in R have similar widths and heights and in addition their centroids are aligned along the same line. The x -projection technique projects onto a line all pixels within the region R , column by column (see Fig. 6). The deepest valleys provide important information to segment the characters in R . Let XP_R be the vector that stores the x -projection of R . Then, the deepest valleys of XP_R can be detected by thresholding it. In this way, several characters C_i can

be extracted from R . Let \overline{W}_R and \overline{H}_R be, respectively, the average width and height of all characters $C_i \in R$. Then, the width and height variances can be calculated, respectively, by the following equations:

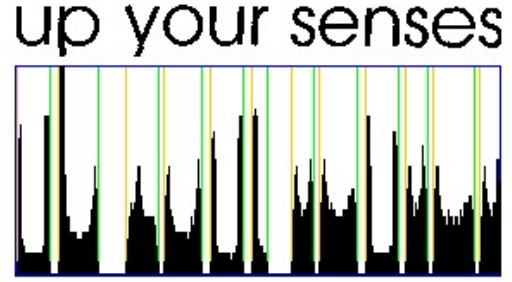


Figure 6: Example of the result by the application of x -projection. Note that yellow and green lines indicate, respectively, the beginning and the end of a character segmentation.

$$\sigma_{W,R}^2 = \frac{1}{N_C} \cdot \sum_{C_i \in R} (W_{C_i} - \overline{W}_R)^2, \quad (5)$$

$$\sigma_{H,R}^2 = \frac{1}{N_C} \cdot \sum_{C_i \in R} (H_{C_i} - \overline{H}_R)^2, \quad (6)$$

where N_C is the number of characters within R . In addition, all character centers form a line and satisfy the equation $y = m_R^C \cdot x + b_R^C$, where m_R^C and b_R^C can be easily obtained by linear regression [16]. Then, the linearity of R can be measured by

$$Lin(R) = \frac{1}{N_C} \cdot \sum_{C_i \in R} \frac{|y_{C_i} - m_R^C \cdot x_{C_i} + b_R^C|}{\sqrt{(m_R^C)^2 + 1}}. \quad (7)$$

Besides considering geometric aspects of R , we also take account that all characters within a text region must have similar gray levels. Let μ be the average gray level of all characters $C_i \in R$ within the image input f . Then, the homogeneity of R can be measured by

$$Hom(R) = \frac{1}{N_P} \cdot \sum_{C_i \in R} \sum_{(x,y) \in C_i} (f_{(x,y)} - \mu)^2 \quad (8)$$

where N_P is the number of pixels of all characters $C_i \in R$. This measure was inspired from the work [9].

Therefore, the considered features to classify text and non-text regions are:

1. *Contrast based feature*
 - (a) Gray level homogeneity (Eq. 8).
2. *Font geometric based features*
 - (a) Number of characters obtained from the x -projection;
 - (b) Height variance of the characters (Eq. 5);

- (c) Width variance of the characters (Eq. 6);
- (d) Linearity of the characters (Eq. 7).

Classifier Design In this work, a decision tree is used to classify $R \in \zeta$ in text or non-text region. For that, a training set $P = \{(p_i, c_i) \in \mathbb{R}^5 \times \{1, 2\} : i = 1, 2, \dots, n\}$ of labeled feature vectors (obtained by the application of the measures described in the previous subsection to text and non-text regions) was built. In Figs. 7(a) and 7(b), we present some examples of text and non-text regions used to build the set P .

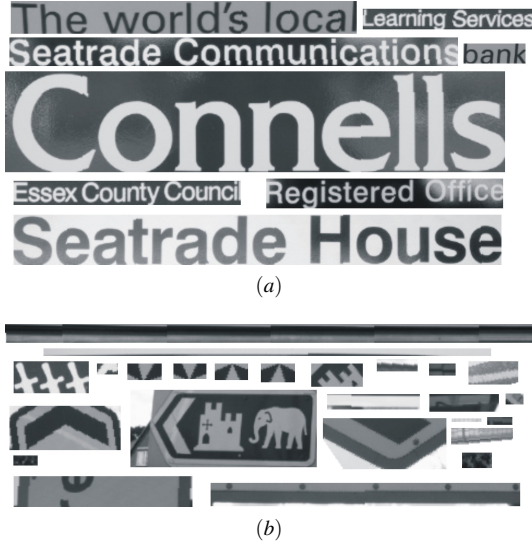


Figure 7: Examples of (a) text regions (class 1) and (b) non-text regions (class 2).

We have used the C4.5 algorithm [17] (implemented in WEKA¹ software) for training the decision tree. The training set P has 683 patterns extracted from 50 images of ICDAR dataset [13] in which 203 patterns belong to class 1 (text region) and 480 patterns to class 2 (non-text region). The traditional cross-validation technique has been used to validate the classifier performance [2]. The training error obtained for the built decision tree is 4%.

4 EXPERIMENTAL RESULTS

In order to evaluate the proposed method, we have selected other 84 images from the ICDAR dataset with different colors, illuminations, scales and orientations as well as images with partial occluded text regions. Like other methods [9, 18, 19], the performance of the proposed method is measured in terms of recall and precision rates.

The 84 test images contain 367 text regions (total text regions) with at least 3 characters. Performed tests showed that 323 text regions were correctly classified

(true positives) and only 7 were wrongly assigned as text regions (false positives). In this case, the recall and precision rates are 88% and 97%, respectively. Table 1 shows the obtained confusion matrix and Fig. 8 presents some images with text regions marked with red rectangles after the application of the decision tree classifier.

Table 1: Confusion matrix.

Classes	True positive	False positive
Text region	323(88%)	7(1.9%)
Non-text region	(98%)	44(12%)

Since our method builds the set ζ of the potential text regions (see Section 3.2) and this set may not contain all the actual text regions within the input image, we can evaluate the proposed method by calculating the recall and precision rates considering only the text regions in ζ . In our experiment, the cardinality of ζ is 347. This means that 20 text regions of the input images have not been detected by the selection and merging procedure. Thus, in this case, the recall and precision rates are 93% and 97.8%, respectively. Table 2 presents the confusion matrix considering only the selected regions.

Table 2: Confusion matrix (considering only the selected regions).

Classes	True positive	False positive
Text region	323(93%)	7(2.1%)
Non-text region	(97.9%)	24(7%)

In order to have a comparison between our method and at least another one, we have applied the Wu's method [19] to the same image dataset. Performed tests showed that only 280 text regions were correctly classified, corresponding to the rate of 77%. However, the Wu's method had a larger number of false positives (75 regions wrongly assigned as text regions) corresponding to the rate of 20%. This happens mainly because Wu's method uses simple thresholds (constant values) in several heuristics for text region classification and consequently it is not sufficiently robust to deal with a large number of images containing text regions with different scales and orientations. Table 3 presents the confusion matrix obtained by Wu's method [19].

Table 3: Confusion matrix of method [19].

Classes	True positive	False positive
Text region	280(77%)	75(20%)
Non-text region	(80%)	87(23%)

We should remark that the performance of Wu's method [19] can be drastically affected in the feature extraction stage where it would have a big difficult to

¹ <http://www.cs.waikato.ac.nz/ml/weka/>



Figure 8: Scene images. (a – f) The classified text regions are marked by red rectangles.

extract text regions containing characters that (1) are not horizontally aligned along a line or (2) have different sizes. This is due to the fact that in this stage it uses a fixed (1×7) horizontal line as a SE to perform two morphological operators: an opening and a closing.

We should remark that the performance of Wu’s method [19] can be drastically affected in the feature extraction stage, since, at this stage, it uses a fixed (1×7) horizontal line as a SE to perform two morphological operators: an opening and a closing. Consequently, it would have a big difficult to extract

text regions containing characters that (1) are not horizontally aligned along a line or (2) have different sizes. Just for a quick comparison, Fig. 9(a) shows an application of Wu’s method [19] to a scene image; while in Fig. 9(b), we show the result for the same image using our method.

5 CONCLUSION AND FUTURE WORK

We presented a new method to classify regions extracted from scene images by morphological filters in



(a)



(b)

Figure 9: Scene images containing text regions with different scales. Result from the application of (a) Wu's method [19] and (b) our method.

text or non-text region using a decision tree. Our technique can be divided into three parts. Firstly, we extract a set of regions by a robust scheme based on morphological filters. Then, after a refinement, a set of text attributes is obtained for each region. In the last step, a decision tree is built in order to classify them as text or non-text regions. The obtained results show a good performance of text region classification with the overall recall and precision rates equal to 88% and 97%, respectively. These results show that our method can be a better alternative for text localization in scene images. For future work, we envisage the following points: (i) perform a comparative analysis of our method with others using the metric proposed in [14]; (ii) propose a scheme for extracting text region candidates with different character sizes presented in the input image using the ultimate closing and opening operators [18]; (iii) propose an algorithm to adjust the exact location of the classified text region in the input image.

ACKNOWLEDGEMENTS

We would like to thank the financial support from CNPq (www.cnpq.br) and FAPESP (www.fapesp.br).

REFERENCES

[1] Rodolfo P. dos Santos, Gabriela S. Clemente, Tsang Ing Ren, and George D.C. Cavalcanti. Text line segmentation based on morphology and histogram projection. *Document Analysis and Recognition, International Conference on*, 0:651–655, 2009.

[2] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern Classification*. Wiley-Interscience, second edition, 2000.

[3] Sergio Escalera, Xavier Baró, Jordi Vitrià, and Petia Radeva. Text detection in urban scenes. *Catalan Association of Artificial Intelligence, 12th International Congress of the*, 2009.

[4] Julinda Gllavata, Ermir Qeli, and Bernd Freisleben. Detecting text in videos using fuzzy clustering ensembles. *Multimedia, International Symposium on*, 0:283–290, 2006.

[5] Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing (3rd Edition)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2006.

[6] Lixu Gu. Character region identification from cover images using dt. In *ICADL*, pages 388–397, 2004.

[7] Lixu Gu, Toyohisa Kaneko, Naoki Tanaka, and R M Haralick. Robust extraction of characters from color scene image using mathematical morphology. *Pattern Recognition, International Conference on*, 2:1002, 1998.

[8] Shehzad Muhammad Hanif and Lionel Prevost. Text detection and localization in complex scene images using constrained adaboost algorithm. *Document Analysis and Recognition, International Conference on*, 0:1–5, 2009.

[9] Y. M. Y. Hasan and L. J. Karam. Morphological text extraction from images. *Image Processing, IEEE Transactions on*, 9(11):1978–1983, 2000.

[10] Renjie Jiang, Feihu Qi, Li Xu, and Guorong Wu. Detecting and segmenting text from natural scenes with 2-stage classification. In *ISDA '06: Proceedings of the Sixth International Conference on Intelligent Systems Design and Applications*, pages 819–824, Washington, DC, USA, 2006.

[11] Keechul Jung, Kwang In Kim, and Anil K. Jain. Text information extraction in images and video: a survey. *Pattern Recognition*, 37(5):977–997, 2004.

[12] Kwang In Kim, Keechul Jung, and Jin Hyung Kim. Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(12):1631–1639, 2003.

[13] S. M. Lucas, A. Panaretos, L. Sosa, A. Tang, S. Wong, and R. Young. Icdar 2003 robust reading competitions. *Document Analysis and Recognition, Int. Conf. on*, 2:682, 2003.

[14] 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. *International Journal on Document Analysis and Recognition - Special Issue on Camera-based Text and Document Recognition*, 7(2-3):105–122, 2005.

[15] J. R. Parker and P. Federl. An approach to licence plate recognition. In *Proceedings of Visual Interface*, 1997.

[16] William H. Press, Brian P. Flannery, Saul A. Teukolsky, and William T. Vetterling. *Numerical recipes in C: the art of scientific computing*. Cambridge University Press, New York, NY, USA, 1988.

[17] J. Ross Quinlan. *C4.5: programs for machine learning*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1993.

[18] Thomas Retornaz and Beatriz Marcotegui. Scene text localization based on the ultimate opening. In *Proceedings*, volume 1, pages 177–188. Instituto Nacional de Pesquisas Espaciais (INPE), 2007.

[19] Jui-Chen Wu, Jun-Wei Hsieh, and Yung-Sheng Chen. Morphology-based text line extraction. *Machine Vision Application*, 19(3):195–207, 2008.