

LOW-LEVEL FEATURE'S SET FOR TEXT IMAGE DISCRIMINATION

N. G. Alessi*, S. Battiato**, G. Gallo*, M. Mancuso**, and F. Stanco*
* Dipartimento di Matematica e Informatica - Viale A. Doria, 6 - 95125 Catania (Italy)
** STMicroelectronics - Stradale Primosole, 50 - 95121 Catania (Italy)

ABSTRACT

Automatic discrimination of digital images based on their semantic content is of great relevance in daily application. In this paper it is proposed a system that allows to detect if a digital image contains a text document. Our technique is a multi-steps procedure based on low-level feature's set computation. The experimental results show that the proposed algorithm is competitive in efficiency with classical techniques, with a lower complexity.

1. INTRODUCTION

In this paper we address the problem of text images discrimination, (i.e. to decide if an image contains text or not). There are several methods of semantic classification for images relative to text segmentation ([2], [8], [10], [11], [14], [15], [19]), or relative to other properties ([9], [16], [18]). Since text/non-text classification could be applied in real-time applications related to digital photography ([4], [5]), the reduction of processing time is a relevant issue. This is instrumental for image capturing devices that focus on quality enhancement and reduction of processing time ([3], [6], [7]).

General classification methods extract a representative set of low-level features from the input image ([1], [13], [17]) and hence determine the relative semantic class. Classifying text images in such a way would be easier in case of scanned images because they present an uniform background and constant illumination properties. Unfortunately, when the input image is obtained by a consumer device (e.g. mobile phone, digital still camera [4]) most of the low-level features that characterize a text image become unusable, mainly due to the particular illumination conditions that affect the image background (Figure 1). In this case, to be able to discriminate correctly between text/non-text images is useful to apply specific techniques of enhancement/compression provided by the particular *Image Generation Pipeline* ([4], [12]) used in the generation process.

The proposed approach splits the input image into regular squares. Each square in turn is classified as text/non-text. Statistics relative to this classification are eventually used to classify the whole document. The algorithm works like a segmentation method, but it is more selective to avoid false positives.

The rest of the paper is organized in the following way. Section 1 summarizes the underlying criteria used in text/non-text discrimination reporting in detail all steps

needed to identify the text areas of input images. In Section 2 the experiments performed on images acquired by real CCD/CMOS sensors are briefly reported. A final Section closes the paper tracking directions for future works.

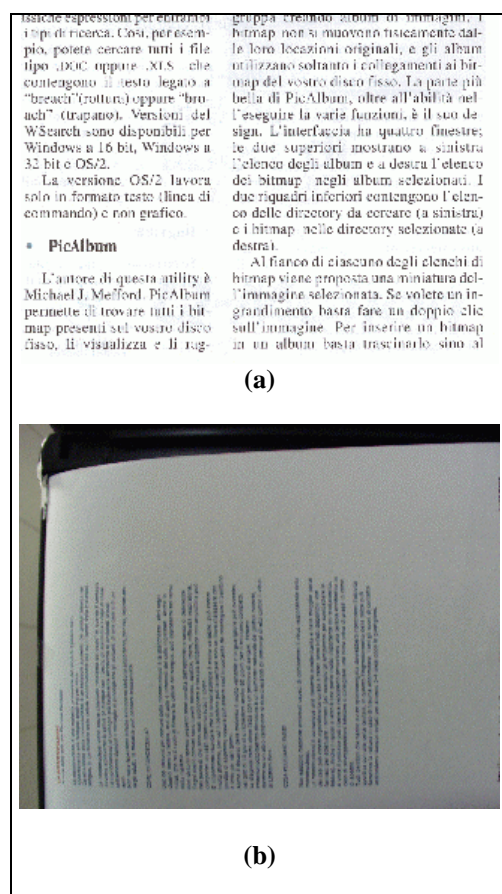


Figure 1 - (a) A scanned text image; (b) A text image acquired by a digital still camera.

2. THE PROPOSED APPROACH

The proposed method works analyzing a set of low-level features over regular blocks extracted from the image. The feature set has to be chosen properly in order to identify text images avoiding false positives (i.e. the feature set should minimize the probability to classify as text a non text area). Classification of the whole image is based on the evidences collected on this blocks. Typical text areas show:

1. Fast changes in intensity (e.g. high values of gradient magnitude).
2. Luminance changes tend to be very frequent and without a preferred direction.
3. In a sufficiently wide text area there is a connected zone with an almost constant value of luminance (the background).

Only if these requisites are satisfied the method gives a positive answer about the text presence in the area. The algorithm can be decomposed in practical steps according to the aforementioned rules.

Step 1. The first discrete derivative is computed for each pixel of the input images. Gradient's magnitude and direction are computed and stored in a matrix G .

Step 2. The matrix G is properly binarized by a thresholding operator as suggested in [15]. The threshold is determined according to the formula:

$$th = \sqrt{4 \frac{\sum_{i=1}^{h-1} \sum_{j=1}^{w-1} G^2(i, j)}{(h-1)(w-1)}} \quad (1)$$

where h , w are respectively the height and width of the image.

The set of remaining pixels E contains only the pixels satisfying the above assumption 1. Starting from the input image showed in Fig. 2-a, the graphical representation of the corresponding matrix G together with its binarization, are reported in Fig. 2-b and 2-c.

Step 3. Discard from E all the pixels not corresponding to local maxima in G . This step, needed to satisfy the second assumption, is carried out performing a simplified *non-maxima suppression*. It determines, for each pixel, the direction of the gradient vector and then interpolates its magnitude along the obtained direction in a suitable neighborhood (typically 3×3); the pixel is not discarded only if its associated gradient magnitude is not lower than the interpolated ones. Figure 2-d shows the matrix E after this kind of processing.

Step 4. In this step all the boundaries (e.g. horizontal/vertical lines, boundarie, etc...) that have a high pixel's *degree of impurity* M (see [8] for more details) are discarded. M is computed using the following formula:

$$M = \sum_{q=0}^{2p} (A(\mathbf{q}) - \bar{A})^2 \quad (2)$$

where $A(\mathbf{q})$ is the sum of all the pixels in a neighbor of radius r whose magnitude is in $[\mathbf{q}, \mathbf{q} + \mathbf{p}/8]$ and \bar{A} is the average magnitude. This measure is applied over all the pixels that have value greater than a suitable threshold t . In these points the gradient is higher than the local maxima; it's no rare that such pixels are not discarded during the previous steps. Figure 2-e shows the remaining data after the boundaries elimination process.

Step 5. The current binary matrix is divided into small squares of size $b \times b$. Each block not containing at least a

percentual p of text pixels is discarded. Both b and p values are chosen depending on the input image resolution size.

Step 6. To each block of the original input image, identified in the previous step, is applied a requantization in a smaller number of gray levels (typically 16).

Step 7. Finally the *Color Coherence Vector (CCV)* [13] computation is performed on these new blocks to check the presence of a predominant zone of the same quantized graylevel. This step needs of two parameters: the coherence threshold τ , and the percentage of predominant color. Both are introduced by the user. Figure 2-f shows the remaining blocks (i.e. after the background uniformity test).

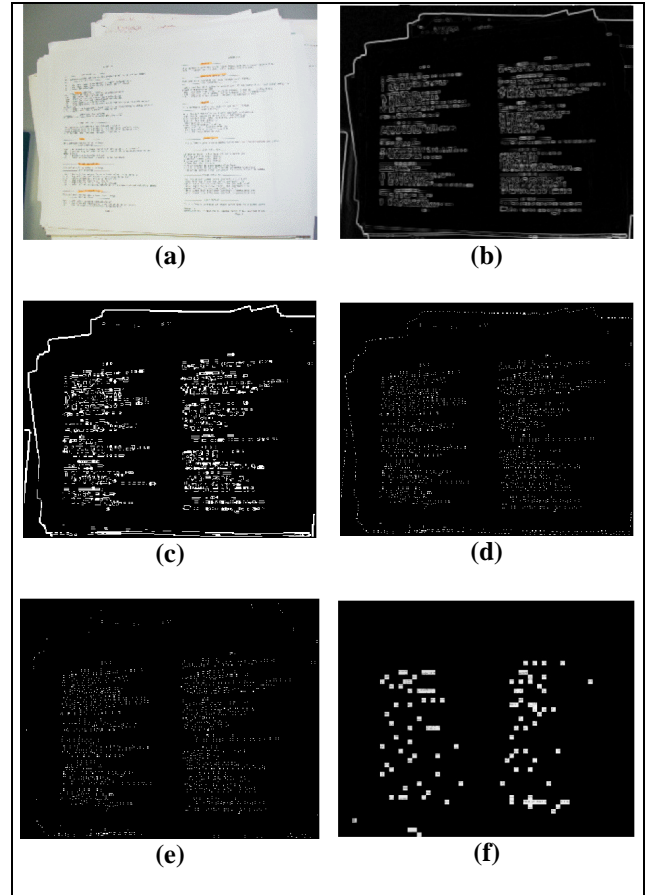


Figure 2 - The algorithm's steps: (a) original image; (b) Sobel gradient magnitude; (c) after thresholding; (d) non-maxima suppression; (e) boundaries elimination; (f) final result.

The output of the overall pipeline is a list of blocks belonging to the original image marked as text blocks. The ratio between the number of total blocks of the image and the number of text blocks gives the *text-blocks* ratio. If it is high then the image contains text, otherwise the picture does not contain a relevant text area. Because of the strictness of the employed filters, the percentage of text blocks for text images computed with the above pipeline will be lower than the real one. However in non-text images this number is negligible enabling us to realize an easy discrimination between the considered semantic classes.

3. EXPERIMENTAL RESULTS

The proposed method has been tested on a large image database with text and non-text images at two different resolutions (640x480, 320x200) containing image belonging to different semantic categories (90 text images and 125 non-text images). The images have been acquired through a digital still camera under several lighting conditions, using a scanner and from clipart CDs. Into both semantic classes (Text, Non-Text) were intentionally inserted images whose semantic class is ambiguous and difficult to verify. The algorithm has been implemented in ANSI C and tested on Intel architecture (366 Mhz Pentium II processor) under the Windows 98 operating system. The times reported to analyze a single image are about 1 second on a 640x480 image and 0.37 seconds on a 320x200 one.

Resolution	640x480	320x200
Block size b	11x11	5x5
Minimum text ratio p	5.5%	15%
Coherence threshold t	30%	16%
Colour predominance	45%	48%
M threshold t	0.38	0.38
M radius r	3	2
Re-quantization bitshift	4	4

Table 1 – List of parameters with their corresponding values used in the experiments.

Table 1 reports the parameters values used in the experiments; all of them are derived empirically. Figure 3 shows the *text-blocks* ratio of the images sets. It's possible to observe a clear separation between the values assigned to the two semantic categories of text and non-text. This is realized fixing a suitable threshold, and considering as textual those images whose value is above this threshold.

Image Resolution	Threshold	Affirmative Text Answers	Affirmative Non-text Answers
640x480	0.01	87/90	119/125
320x200	0.0025	85/90	117/125

Table 2 – Results of text/non-Text discrimination.

Table 2 reports thresholds employed and the overall results obtained using the parameters listed in table 1. The method is able to correctly classify text/non-text in almost all cases. The overall computation time needed to apply the proposed pipeline is negligible and well suited for real time applications.

Figure 4 shows two cases of misclassification for images taken respectively from texts and non-text. The main causes of misclassification on text are due to bad lighting conditions and to noisy backgrounds that causes the final uniformity test to fail. False positives are due to particular (and quite rare) color conditions that result in losing high frequency content during the final requantization.

4. CONCLUSIONS

An algorithm able to discriminate between text/non-text images has been proposed; the relative technique extracts a set of low-level features properly thresholding the final *text-blocks* ratio. Experimental results show that the overall computation is *light* and robust.

Future works will include the generalization to Bayer data images [4], acquired by digital acquisition devices.

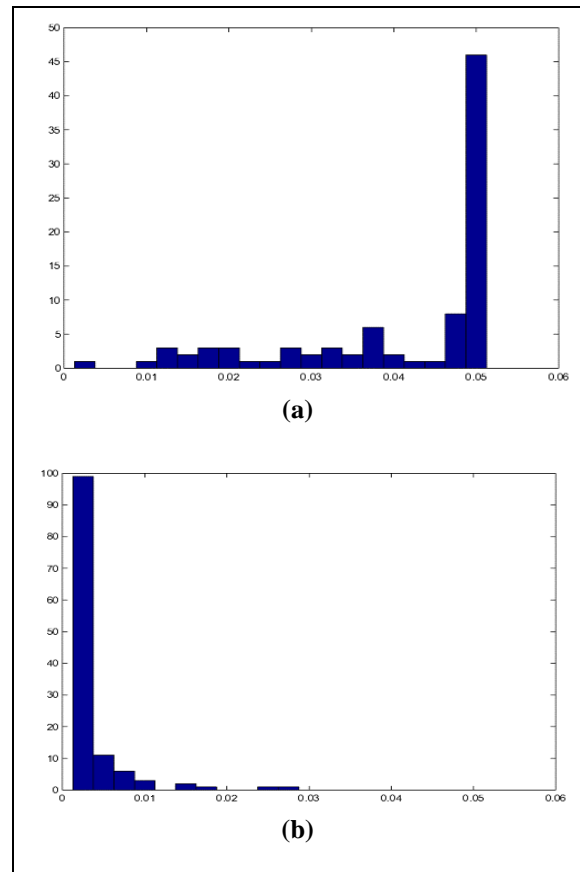


Figure 3 - Experimental results used to properly choose the final threshold value: x-axis reports the value returned by the algorithm; y-axis reports the number of images for a given value. (a) Text; (b) Non-text.

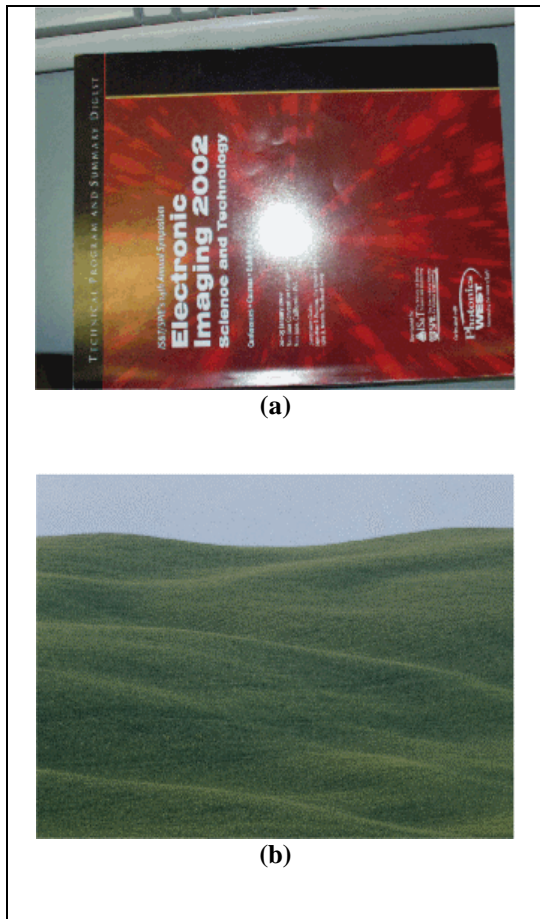


Figure 4 - Examples of misclassified text/non-text images:
 (a) original text image; (b) original non-text image;

REFERENCES

- [1] G. Amadasun, R. King, "Textural features corresponding to textural properties", *IEEE Transaction on System, Man and Cybernetics*, 19(5): 1264-1274, 1989.
- [2] A. Antonacopoulos, T. R. Ritchings, "Representation and classification of complex-shaped printed regions using white tiles", *Proceedings of the 3rd ICDAR*, pp. 1132-1135, Montreal, Canada, 1995.
- [3] S. Battiato, A. Castorina, M. Mancuso - "High Dynamic Range Imaging: Overview and Application" - Accepted for publication: *SPIE Journal of Electronic Imaging* - 2002;
- [4] S. Battiato, M. Mancuso - "An Introduction to the Digital Still Camera Technology" - *ST Journal of System Research - Special Issue on Image Processing for Digital Still Camera*, Vol. 2, No.2, December 2001;
- [5] S. Battiato, M. Mancuso, A. Bosco, M. Guarnera - "Psychovisual and Statistical Optimization of Quantization Tables for DCT Compression Engines" - *In Proceedings of IEEE International Conference on Image Analysis and Processing ICIAP 2001* - pp. 602-606 - Palermo, Italy, September 2001;
- [6] A. Bosco, M. Mancuso, S. Battiato, G. Spampinato - "Adaptive Temporal Filtering for CFA Video Sequences" - *In Proceedings of IEEE ACIVS'02 Advanced Concepts for Intelligent Vision Systems 2002* - pp. 19-24 - Ghent University, Belgium, September 2002;
- [7] A. Bosco, M. Mancuso, S. Battiato, G. Spampinato - "Temporal Noise Reduction of Bayer Matrixed Video Data" - *In Proceedings of IEEE ICME'02 International Conference on Multimedia and Expo 2002* - pp.681-684 - Lausanne, Switzerland, August 2002;
- [8] P. Clark, M. Mirmehdi, "Finding text regions using localized measures", *International Conference on Pattern Recognition ICPR00*, Barcelona, Spain, 2000.
- [9] M. De Ponti, R. Schettini, C. Brambilla, A. Valsasna, "Content-based classification of colour images", *ST Journal of System Research*, Vol.1. No.1, 2001.
- [10] A. K. Jain, S. Bhattacharjee, "Page segmentation using Gabor filters for automatic document processing", *Machine Vision and Applications*, 5(3), pp.169-184, 1992.
- [11] A. K. Jain, Y. Zhong, "Page segmentation using texture analysis", *Pattern Recognition*, 29(5), pp.743-770, 1996.
- [12] G. Messina, S. Battiato, M. Mancuso, A. Buemi - "Improving Image Resolution by Adaptive Back-projection Correction Techniques" - *IEEE Transaction on Consumer Electronics* - Vol. 48, Issue 3, pp. 409-416, August 2002;
- [13] G. Pass, R. Zabih, J. Miller, "Comparing images using color coherence vectors", *Proceedings of Fourth ACM Conference on Multimedia*, (Boston, MA), November 1996.
- [14] J. S. Payne, T. J. Stonham, D. Patel, "Document segmentation using texture analysis", *Proceedings of the 12th ICPR*, pp. 380-382, Jerusalem, Israel, 1994.
- [15] M. Pietikäinen, O. Okun, "Edge-Based Method for Text Detection from Complex Document Images", *Proceedings of the 6th International Conference on Document Analysis and Recognition*, Seattle, USA, pp. 286-291, Sept. 2001.
- [16] M. Szummer, R. Picard, "Indoor-Outdoor Image Classification", *IEEE International Workshop on Content-based Access of Image and Video Databases, in conjunction with ICCV '98*, (Bombay, India), January 1998.
- [17] H. Tamura, S. Mori, T. Yamawaki, "Textural features corresponding to visual perception", *IEEE Transaction on System, Man and Cybernetics*, 8: 460-473, 1978.
- [18] A. Vailaya, A. K. Jain, H. J. Zhang, "On Image Classification: City Images vs. Landscapes", *Pattern Recognition*, vol.31, no.12, pp.1921-1936, 1998.
- [19] D. Wang, S. N. Srihari, "Classification of newspaper image blocks using texture analysis", *Computer vision, Graphics, and Image Processing*, 47, pp. 327-352, 1989.