

SWT voting-based color reduction method for detecting text in natural scene images

Andrej Ikica

Computer Vision Laboratory, Faculty of Computer and Information Science, Tržaška 25, Ljubljana, Slovenia

Advisor/s: Peter Peer

Date and location of the PhD thesis defense: 24 October 2013, University of Ljubljana

Received 28th Jan 2014; accepted 26th May 2014

1 Abstract

In our PhD thesis [1] we give a very detailed and in-depth survey of natural scene text detection methods and propose two novel methods, namely SWT (Stroke Width Transform) voting-based color reduction method [2] and SWT direction determination method [2]. SWT voting-based color reduction method (to which we will refer also as SWT-V) is a novel text detection method that - opposed to many other text detection methods - combines both structural and color information in order to detect text. The proposed method upgrades the text detection oriented color reduction method (to which we will refer to as TOCR) [3] with the additional SWT voting stage and substantially outperforms other state-of-the-art text detection methods. All the image colors rich with SWT pixels [4] that most likely belong to text characters are blocked from being mean-shifted away in the color reduction process. One of the disadvantages of the SWT method [4], however, is the problem of 'light text on the dark background' described in the following paragraphs. To cope with the problem and in order to provide true SWT values to the SWT voting stage we propose an adaptive SWT direction determination method. The method uses SWT profiles to partition an image into subblocks and analyses their SWT histograms of both SWT search directions [2]. Text detection literature does not explicitly address the SWT direction issue, therefore, the proposed method represents a unique scientific contribution to the research field. All text detection methods were evaluated on the CVL OCR DB text detection evaluation dataset [2, 5].

The main disadvantage of the TOCR method [3] stems from the fact that non-dominant text colors often vanish in the color reduction stage. When a certain text color lies too close to a more dominant background color in the RGB color space it is mean-shifted away from its original position. A typical example of vanishing text color is depicted in Figure 1b. A blue 'Clairefontaine' color in Figure 1a is lost in the color reduction stage. To avoid such a problematic behavior we propose the SWT voting-based color reduction method. The proposed method upgrades the TOCR method with the additional stages (depicted as yellow rectangles in Figure 2), namely 'SWT filter', 'SWT direction determination' and 'SWT voting'.

In the 'SWT filter' stage we create a SWT image [4] and obtain its correct SWT directions in the 'SWT direction determination' stage. Since the non-zero pixels in the SWT image most likely belong to the text regions in the image [2] they serve as a very useful clue in the subsequent 'SWT voting' stage.

Correspondence to: andrej.ikica@fri.uni-lj.si

Recommended for acceptance by <Alicia Fornés and Volkmar Frinken>

ELCVIA ISSN: 1577-5097

Published by Computer Vision Center / Universitat Autònoma de Barcelona, Barcelona, Spain

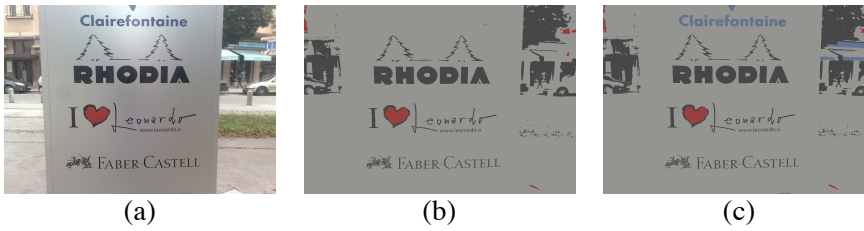


Figure 1: (a) Original image. (b) TOCR result. (c) SWT-V result.

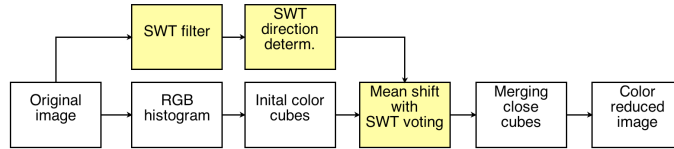


Figure 2: Flowchart of the SWT-V method.

corresponding SWT values in the SWT image. Before each mean-shift step a particular color is looked up in the corresponding SWT lookup table and classified as a text/non-text color using linear discriminant function. All colors classified as text colors are blocked from being mean-shifted any further. The following features are used for classification: SWT density D_{SWT} , standard deviation of SWT lengths SD_{SWTL} and standard deviation of SWT offsets SD_{SWTO} . An example of correctly preserved blue ‘Clairefontaine’ color with the SWT-V method is depicted in Figure 1c.

We evaluated the proposed SWT-V text detection method as well as the TOCR and SWT text detection methods on the CVL OCR DB text detection evaluation dataset [5]. The results are depicted in Table 1. A detailed description of the evaluation scheme is depicted in [2]. Clearly, SWT-V method substantially outperforms both TOCR and SWT methods at all thresholds T_R .

Method	CC error rate (%)					Mean error rate (%)				
	T_R					T_R				
	0.5	0.6	0.7	0.8	0.9	0.5	0.6	0.7	0.8	0.9
TOCR	69.02	65.75	61.86	56.95	48.53	77.71	74.40	71.06	66.02	57.40
SWT-V	71.01	67.48	63.93	59.84	50.41	80.24	76.72	73.77	69.12	59.27
SWT	61.76	54.26	43.34	28.81	9.72	68.47	62.27	53.38	36.78	14.87

Table 1: Text detection results.

The SWT voting concept substantially improves the color reduction process. The SWT voting-based color reduction method clearly outperforms both TOCR and SWT on the CVL OCR DB evaluation dataset. A very important aspect of our work, as well, is a SWT direction determination method, which is able to find correct SWT values when both light text on dark background and dark text on light background appear in the image at the same time. The SWT direction determination method represents an important scientific contribution to the research field.

References

- [1] A. Ikica, “Text detection methods in images of natural scenes”, *Ph.D. Thesis*, University of Ljubljana, Slovenia, 2013. (in Slovenian)
- [2] A. Ikica, P. Peer, “Swt voting-based color reduction for text detection in natural scene images”, *EURASIP Journal on Advances in Signal Processing*, 2013(1):1-13, 2013.
- [3] N. Nikolaou, N. Papamarkos, “Color reduction for complex document images”, *International Journal of Imaging Systems and Technology*, 19(1):14-26, 2009.
- [4] B. Epshtein, E. Ofek, Y. Wexler, “Detecting text in natural scenes with stroke width transform”, *IEEE Proc. of Conf. on Computer Vision and Pattern Recognition*, 2963-2970, 2010.
- [5] A. Ikica, P. Peer, “Cvl ocr db, an annotated image database of text in natural scenes, and its usability”, *Info. MIDEM*, 41(2):150-154, 2011.
- [6] C. Yi, Y.L. Tian, “Text string detection from natural scenes by structure-based partition and grouping”, *IEEE Transactions on Image Processing*, 20(9):2594-2605, 2011.