International Science Group

ISG-KONF.COM

IV

INTERNATIONAL SCIENTIFIC AND PRACTICAL CONFERENCE "SCIENCE, THEORY AND PRACTICE"

Tokyo, Japan
October 12-15, 2021

ISBN 978-1-63972-064-4

DOI 10.46299/ISG.2021.II.IV

SCIENCE, THEORY AND PRACTICE

Abstracts of IV International Scientific and Practical Conference

Tokyo, Japan
October 12 – 15, 2021

Library of Congress Cataloging-in-Publication Data

UDC 01.1
The IV International Science Conference «Science, theory and practice», October 12 – 15, 2021, Tokyo, Japan. 477 p.

ISBN - 978-1-63972-064-4
DOI - 10.46299/ISG.2021.II.IV

EDITORIAL BOARD

RESEARCH AND IMPLEMENTATION OF THE VIDEO STREAM TEXT RECOGNITION METHOD

Vechirska Iryna,
Doctor of Technical Sciences, professor
Kharkiv National University of Radio Electronics

Tymofieiev Oleksii,
Master in informatics
Kharkiv National University of Radio Electronics

One of the major design goals of our method was the ability to extract text from still images as well as video sequences. Therefore, in the case of video sequences, Spatio-temporal approaches to text detection were not an option, which treat the 3D input stream as such and detect the text directly in the 3D space. Instead, a tracking method treats the input stream as a temporal sequence of images and detects text on a frame by frame basis [1-3]. The detected text objects are tracked across several frames in order to create the text appearance, which is a result of the detection and the tracking process. In this article, we concentrate on static, non-moving text. A generalization of our work to simple, linear movement has been done by Marquis.

Apart from the detection itself, the additional temporal component in video sequences also leads to consequences and open questions concerning the usage of the detected appearance. The detected text forms a 3D object [4] with a temporal dimension, which cannot be recognized by OCR software directly, at least not by existing OCR technology (Figure 1). Several possible techniques may be considered:

– Choose a single frame out of all possible frames of the appearance. This is a technique, which has the disadvantage of losing the possible additional temporal information in the appearance;

– Use the 3D text object as it is and develops a new recognition algorithm which exploits the temporal information directly in order to increase recognition performance;

– Integrate the 3D appearance into a single 2D image, exploiting the temporal information in order to “clean up” the image and to create a single image of better quality;

– Apply the recognition algorithm to each of the frames in order to create a set of recognized text strings;

– Apply symbolic statistics to the set in order to create a single string with the most probable contents.

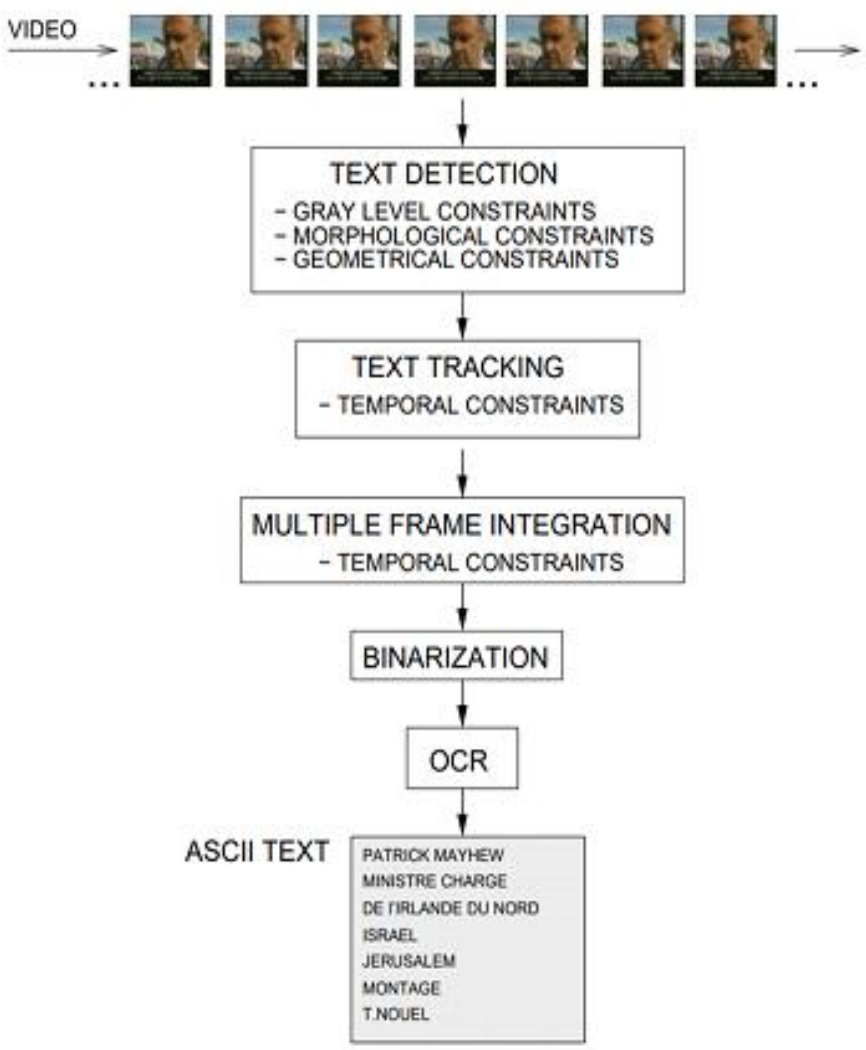

Figure 1. The scheme of our system

The first solution – choosing the text rectangle from a single frame for recognition – fails due do the miserable quality of most text taken from videos. At least some processing is necessary to enhance the image quality.

The second solution is beyond the scope of this work. The development of a new OCR technology needs a tremendous amount of engineering experience in order to find the necessary heuristics which allow these systems to obtain their excellent recognition performance. Instead, we apply commercial software, which delivers excellent results on scanned printed or faxed documents.

Unfortunately, commercial OCR software is not adapted to the type of data, which is why we chose the third solution: We integrate the text appearance into a single image of better quality, which is closer to the type of data expected by commercial OCR software.

Our research team also successfully worked on the fourth way to use the text appearance. An algorithm which uses new theoretical research on statistical processing of character strings done by our team is given in [5-7].

As already stated, the detection algorithm for still images is applied to each frame of the sequence separately [8]. The detected text rectangles are passed to a tracking step, which finds corresponding rectangles of the same text appearance in different

frames. From several frames of an appearance, a single enhanced image is generated and binarized, i.e. segmented into characters and background, before passing it to standard commercial OCR software. Text in videos has gray level properties (e.g. high contrast in given directions), morphological properties (spatial distribution, shape), geometrical properties (length, ratio height/length etc.) and temporal properties (stability) [9]. Our method makes use of these properties, starting from the signal and going sequentially to the more domain dependent properties. The final step (the character segmentation) results in a set of binary boxes containing text which need to be recognized by a classical commercial OCR system.

References:

1. Kobylin O., Gorokhovatskyi V., Tvoroshenko I., and Peredrii O. (2020) The application of non-parametric statistics methods in image classifiers based on structural description components, Telecommunications and Radio Engineering, 79(10), pp. 855-863.

2. Кобилін О.А., Творошенко І.С. (2021) Методи цифрової обробки зображень: навч. посібник. Харків: ХНУРЕ, 124 с.

3. Творошенко І.С. (2021) Технології прийняття рішень в інформаційних системах: навч. посібник. Харків: ХНУРЕ, 120 с.

4. M. Ayaz Ahmad, Irina Tvoroshenko, Jalal Hasan Baker, Liubov Kochura, Vyacheslav Lyashenko (2020) Interactive Geoinformation Three-Dimensional Model of a Landscape Park Using Geoinformatics Tools, International Journal on Advanced Science, Engineering and Information Technology, 10(5), pp. 2005-2013.

5. Гороховатский В.А., Передрий Е.О. (2009) Корреляционные методы распознавания изображений путем голосования систем фрагментов, *Радіоелектроніка. Інформатика. Управління*, 1(20), С. 74-81.

6. Gadetska S.V., and Gorokhovatsky V.A. (2018) Statistical Measures for Computation of the Image Relevance of Visual Objects in the Structural Image Classification Methods, *Telecommunications and Radio Engineering*, 77(12), pp. 1041-1053.

7. Gorokhovatskiy V.A. (2011) Compression of Descriptions in the Structural Image Recognition, *Telecommunications and Radio Engineering*, 70(15), pp. 1363-1371.

8. M. Ayaz Ahmad, Volodymyr Gorokhovatskyi, Iryna Tvoroshenko, Nataliia Vlasenko, Syed Khalid Mustafa (2021) The Research of Image Classification Methods Based on the Introducing Cluster Representation Parameters for the Structural Description, International Journal of Engineering Trends and Technology, 69(10), pp. 186-192.

9. Daradkeh Y.I., Gorokhovatskyi V., Tvoroshenko I., Gadetska S., and Al-Dhaifallah M. (2021) Methods of Classification of Images on the Basis of the Values of Statistical Distributions for the Composition of Structural Description Components, IEEE Access, 9, pp. 92964-92973.