IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 7, July 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.542

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907172 |

A Unified Framework for Tracking Based Text Detection and Recognition from Web Videos

D Yuvarani, Mrs. Hemavathy M.Tech,

B. Tech Student, Department of Information Technology, Panimalar Engineering College, Chennai, India

Assistant Professor, Department of Information Technology, Panimalar Engineering College, Chennai, India

ABSTRACT: Intelligent analysis of video data is currently in wide demand because video is a major source of sensory data in our lives. Text is a prominent and direct source of information in video, while recent surveys of text detection and recognition in imagery focus mainly on text extraction from scene images. Here, this paper presents a comprehensive survey of text detection, tracking and recognition in video with three major contributions. First, a generic framework is proposed for video text extraction that uniformly describes detection, tracking, recognition, and their relations and interactions. Second, within this framework, a variety of methods, systems and evaluation protocols of video text extraction are summarized, compared, and analyzed. Existing text tracking techniques, tracking based detection and recognition techniques are specifically highlighted. Third, related applications, prominent challenges, and future directions for video text extraction (especially from scene videos and web videos) are also thoroughly discussed.

I. INTRODUCTION

Some researchers have presented specific frameworks for video text extraction. For example, Antani et al. divided video text extraction into four tasks: detection, localization, segmentation, and recognition. In their system, the tracking stage provides additional input to the spatial- temporal decision fusion for improving localization. Jung et al. summarized the sub problems of a text information extraction system for both images and video into text detection, localization, tracking, extraction and enhancement, and recognition . The video text recognition flowchart designed by Elagouni et al. is similar to that of, but added a correction (post-processing) step with natural language processing. In contrast, in this paper we propose a unified video text extraction framework, where text detection, tracking, and recognition techniques are uniformly described and surveyed.

The unified video text **D**etection, **T**racking and **R**ecognition (**DETR**) framework. Here, Detection is the task of localizing the text in each video frame with bounding boxes. Tracking is the task of main-training the integrity of the text location and tracking text across adjacent frames. Recognition involves segmenting (if necessary) text and recognizing it using Optical Character Recognition (OCR) techniques. Obviously, Recognition is performed on text regions detected from Detection results (Detection-based-Recognition), and Tracking uses the locations identified in the Detection step to track text (Tracking).

A unified frame work for text Detection, Tracking and Recognition in video. This framework uniformly describes detection, tracking, recognition (the three main tasks), and their relations and interactions. The major relations among these main tasks are unified as "detection-based- recognition", "tracking-based-detection" and "tracking- based-recognition". The other relations among these tasks are also named as "refinement-by-recognition (for detection)","refinement-by-recognition(for tracking)" and "tracking-with-detection.

II. LITERATURE SURVEY

2.1 LITERATURE REVIEW

A literature review is a body of text that aims to review the critical points of current knowledge on and/or methodological approaches to a particular topic. It is secondary sources and discuss published information in a particular subject area and sometimes information in a particular subject area within a certain time period. Its ultimate goal is to bring the reader up to date with current literature on a topic and forms the basis for another goal, such as

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907172 |

future research that may be needed in the area and precedes a research proposal and may be just a simple summary of sources. Usually, it has an organizational pattern and combines both summary and synthesis.

A summary is a recap of important information about the source, but a synthesis is a re-organization, reshuffling of information. It might give a new interpretation of old material or combine new with old interpretations or it might trace the intellectual progression of the field, including major debates. Depending on the situation, the literature review may evaluate the sources and advise the reader on the most pertinent or relevant of them. Loan default trends have been long studied from a socio-economic stand point. Most economics surveys believe in empirical modeling of these complex systems in order to be able to predict the loan default rate for a particular individual. The use of machine learning for such tasks is a trend which it is observing now. Some of the surveys to understand the past and present perspective of loan approval or not.

2.1.1 A dataset for Arabic text detection, tracking and recognition in news videos.

Author: Oussama Zayene, Jean Hennebert, Sameh Masmoudi Touj, Rolf Ingold, Najoua Essoukri Ben Amara-2015

Description:

Recently, promising results have been reported on video text detection and recognition. Most of the proposed methods are tested on private datasets with non-uniform evaluation metrics. We report here on the development of a publicly accessible annotated video dataset designed to assess the performance of different artificial Arabic text detection, tracking and recognition systems.
The dataset includes 80 videos (more than 850,000 frames) collected from 4 different Arabic news channels. An attempt was made to ensure maximum diversities of the textual content in terms of size, position and background. This data is accompanied by detailed annotations for each text box. We also present a region-based text detection approach in addition to a set of evaluation protocols on which the performance of different systems can be measured.

2.1.2 Joint Object Detection, Tracking and Recognition with Application to Visually Impaired Navigational Assistance
Authors: Ruxandra Tapu, Bogdan Mocanu,

Description:

In this paper, we introduce the so-called ***DEEP-SEE*** framework that jointly exploits computer vision algorithms and deep convolution neural networks (CNNs) to detect, track and recognize in real time objects encountered during navigation.

A first feature concerns an object detection technique designed to localize both static and dynamic objects without any a priori knowledge about their position, type or shape. The methodological core of the proposed approach relies on a novel object tracking method based on two convolution neural networks trained offline. The key principle consists of alternating between tracking using motion information and predicting the object location in time based on visual similarity. The validation of the tracking technique is performed on standard bench mark VOT datasets, and shows that the proposed approach returns state-of-the-art results while minimizing the computational complexity. Then, the ***DEEP-SEE*** framework is integrated into a novel assistive device, designed to improve cognition of VI people and to increase their safety when navigating in crowded urban scenes. The validation of our assistive device is performed on a video dataset with 30 elements acquired with the help of VI users. The proposed system shows high accuracy (>90%) and robustness (>90%) scores regardless on the scene dynamics.

2.1.3 Scene text detection and recognition: recent advances and future trends.

Authors: Yingying Zhu, Cong Yao, Xiang Bai-2015

Description:

Text, as one of the most influential inventions of humanity, has played an important role in human life, so far from ancient times. The rich and precise information embodied in text is very useful in a wide range of vision-based applications, therefore text detection and recognition in natural scenes have become important and active research

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907172 |

topics in computer vision and document analysis. Especially in recent years, the community has seen a surge of research efforts and substantial progresses in these fields, though a variety of challenges (e.g. noise, blur, distortion, occlusion and variation) still remain. The purposes of this survey are three-fold: 1) introduce up-to-date works, 2) identify state-of the-art algorithms, and 3) predict potential research directions in the future.
Moreover, this paper provides comprehensive links to publicly available resources, including benchmark datasets, source codes, and online demos. In summary, this literature review can serve as a good reference for researchers in the are as of scene text detection and recognition.

2.1.4 Tracking Based Multi-Orientation Scene Text Detection: A Unified Framework With Dynamic Programming

Authors : Chun Yang, Xu-Cheng Yin, Wei-Yi Pei, Shu Tian, Ze-YZuo, Chao Zhu– 2017

Description:
There are a variety of grand challenges for multi-orientation text detection in scene videos, where the typical issues include skew distortion, low contrast, and arbitrary motion. Most conventional video text detection methods using individual frames have limited performance. In this paper, we propose a novel tracking based multi-orientation scene text detection method using multiple frames within a unified framework via dynamic programming. First, a multi information fusion-based multi-orientation text detection method in each frame is proposed to extensively locate possible character candidates and extract text regions with multiple channels and scales. Second, an optimal tracking trajectory is learned and linked globally over consecutive frames by dynamic programming to finally refine the detection results with all detection, recognition, and prediction information. Moreover, the effectiveness of our proposed system is evaluated with the state-of-the-art performances on several public data sets of multi-orientation scene text images and videos, including MSRA-TD500, USTB-SV1K, and ICDAR2015 Scene Videos.

2.1.5 Sensor Tag Detection, Tracking and Recognition for AR Application

Author: Stanislav Mukhametshin, Alisa Makhmutova, IgorAnikin-2019

Description:

Augmented reality (AR) applications are widely used in our world today. It may look at the first glance that AR is primarily used in entertainment and learning applications, but in fact, now big companies and corporations try to integrate AR technologies into their industrial processes. It reduces the number of unnecessary communication and increases the net profit of the company.
Most companies use AR devices for front end tasks and a server side for back end. This approach allows reducing the performance requirements for an AR device by implementing time-consuming operations on the powerful server. However, some operations should be implemented directly on the AR device and we have to implement them rationally in terms of performance. Also we have to make interaction between the AR device and a server more effectively. The most important and frequently used operations in AR applications are related to computer vision: image processing, objects detection, tracking and recognition. Therefore, it is important to compare the performance of these operations on AR devices and suggest those that will be implemented on the server side and on the AR device. This paper focuses on deciding these tasks for an AR client-side application for the Siemens company. This application provides a more effective way for monitoring the equipment at remote facilities. Different approaches for sensor tag detection and tracking on a real time video are considered and evaluated. The way for communication between the AR device and the server side is suggested.

2.2 EXISTING SYSTEM

- Our existing system, text has received increasing attention as a key and direct information sources in video.
- As examples, caption text usually an notates information concerning where and when and the events in video happened or who was involved. Hence, text extraction and analysis in video has attracted considerable attention in multimedia understanding systems.
- Scene text, audio, and visual features to construct their retrieval systems.
- Specifically, some researchers performed investigations of only image retrieval by leveraging both textual image representations.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907172 |

Disadvantages:

- Retrieval of text tracking only in images.
- Low accuracy
- Less speed

2.3 PROPOSED SYSTEM

- Video text extraction plays an important role for multimedia understanding and retrieval.
- We propose a generic Bayesian-based framework of Tracking based Text Detection and Recognition (T2DAR) from web videos for embedded captions.
- Which is composed of three major components, i.e., text tracking, tracking based text detection, and tracking based text recognition.
- In this unified framework, text tracking is first conducted by tracking-by detection.

Advantages:

- Easily tracking and text detection in web videos
- High accuracy compared to existing system
- Less time

III. SYSTEM STUDY

3.1. FEASIBILITY STUDY

The feasibility of the project is analyzed in this phase and business proposal is put forth N with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential. Three key considerations involved in the feasibility analysis are

i. Economical Feasibility
ii. Technical Feasibility
iii. Social Feasibility

3.1.1. Economic Feasibility

This study is carried out to check the economic impact that the system will have on theorganization. The amount of fund that the company can pour into the research and developmentof the system is limited. The expenditures must be justified. Thus, the developed system as wellwithin the budget and this was achieved because most of the technologies used are freelyavailable. Only the customized products had to be purchased.

3.1.2. Technical Feasibility

This study is carried out to check the technical feasibility, that is, the technicalrequirements of the system. Any system developed must not have a high demand on the availabletechnical resources. This will lead to high demands on the available technical resources. Thiswill lead to high demands being placed on the client. The developed system must have a modestrequirement, as only minimal or null changes are required for implementing this system.

3.1.3. Social Feasibility

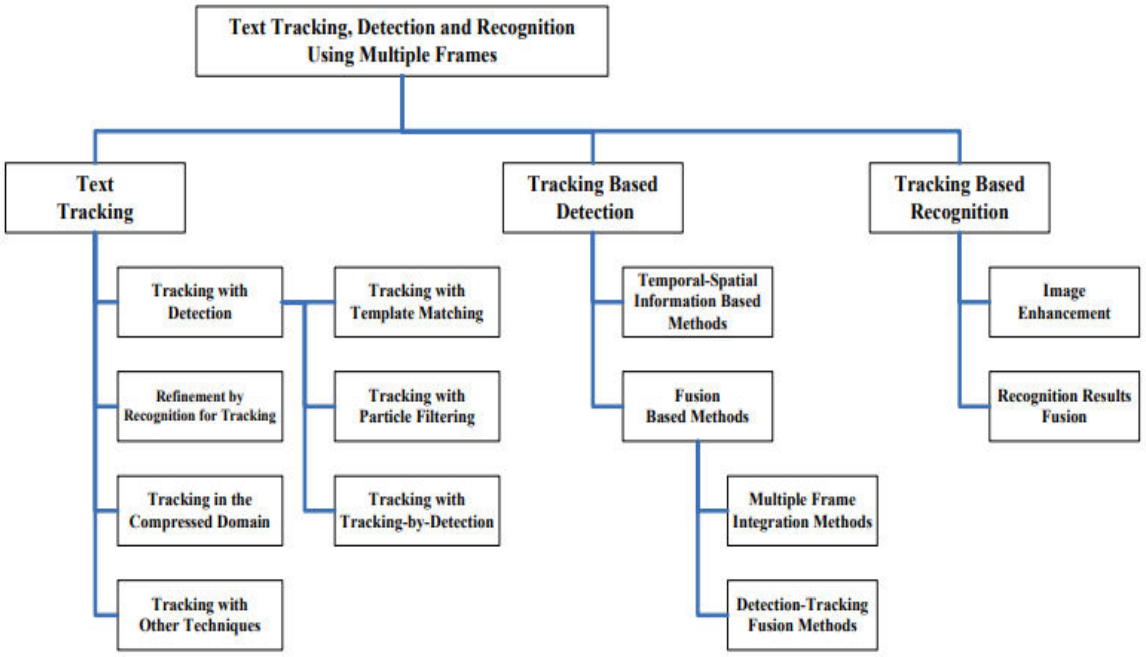
The aspect of study is to check the level of acceptance of the system by the user. Thisincludes the process of training the user to use the system efficiently. The user must not feel
threatened by the system, instead must accept it as a necessity. The level of acceptance by theusers solely depends on the methods that are employed to educate the user about the system andto make him familiar with it. His level of confidence must be raised so that he is also able tomake some constructive criticism, which is welcomed, as he is the final user of the system.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907172 |

IV. SYSTEM ARCHITECTURE



HARDWARE REQUIREMENTS

- CPU type : Intel Pentium 4
- Ram size : 512 MB
- Hard disk capacity : 80 GB
- Monitor type : 15 Inch colour monitor
- Keyboard type : Internet keyboard

V. TESTING

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub – assemblies, assemblies and/or a finished product It is the process of exercising software with the intent of ensuring that the
Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

5.1. TYPES OF TESTS

5.1.1. UNIT TESTING

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application .it is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907172 |

5.1.2. INTEGRATION TESTING

Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfaction, as shown by successfully unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components.

5.1.3. FUNCTIONAL TEST

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals.

Functional testing is centered on the following items:

Valid Input : identified classes of valid input must be accepted.
Invalid Input : identified classes of invalid input must be rejected.
Functions : identified functions must be exercised.
Output : identified classes of application outputs must be exercised
Procedures : interfacing systems or procedures must be invoked.

Organization and preparation of functional tests is focused on requirements, key functions, or special test cases. In addition, systematic coverage pertaining to identify Business process flows; data fields, predefined processes, and successive processes must be considered for testing. Before functional testing is complete, additional tests are identified and the effective value of current tests is determined.

5.1.4. SYSTEM TEST

System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is the configuration oriented system integration test. System testing is based on process descriptions and flows, emphasizing pre-driven process links and integration points.

5.1.5. WHITE BOX TESTING

White Box Testing is a testing in which in which the software tester has knowledge of the inner workings, structure and language of the software, or at least its purpose. It is purpose. It is used to test areas that cannot be reached from a black box level.

5.1.6. BLACK BOX TESTING

Black Box Testing is testing the software without any knowledge of the inner workings, structure or language of the module being tested. Black box tests, as most other kinds of tests, must be written from a definitive source document, such as specification or requirements document, such as specification or requirements document. It is a testing in which the software under test is treated, as a black box .you cannot “see” into it. The test provides inputs and responds to outputs without considering how the software works.

5.1.7. UNIT TESTING:

Unit testing is usually conducted as part of a combined code and unit test phase of the software lifecycle, although it is not uncommon for coding and unit testing to be conducted as two distinct phases.

Test strategy and approach

Field testing will be performed manually and functional tests will be written in detail.

Test objectives

- All field entries must work properly.
- Pages must be activated from the identified link.
- The entry screen, messages and responses must not be delayed.

integration test. System testing is based on process descriptions and flows, emphasizing pre-driven process links and integration points.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907172 |

VI. RESULT AND DISCUSSION

6.1.1 OPEN IN THE ANACONDA NAVIGATOR

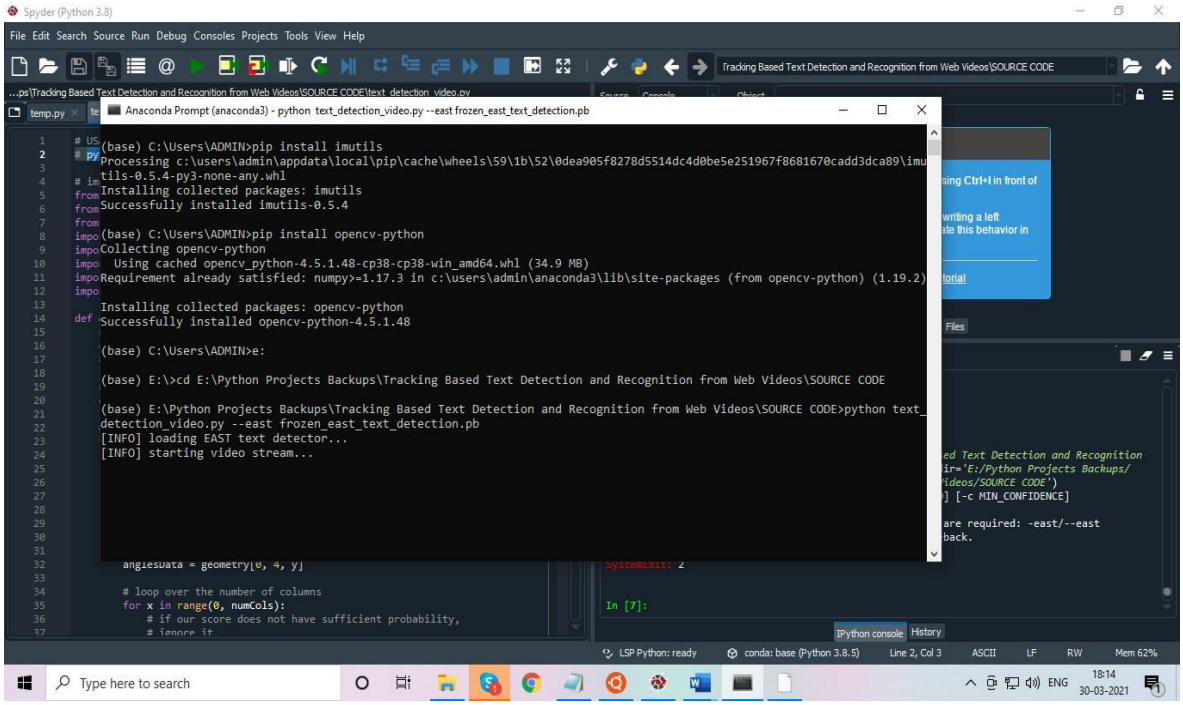


Fig 6.1 : Open In The Anaconda Navigator

6.1.2 LAUNCH THE SPYDER PLATFORM

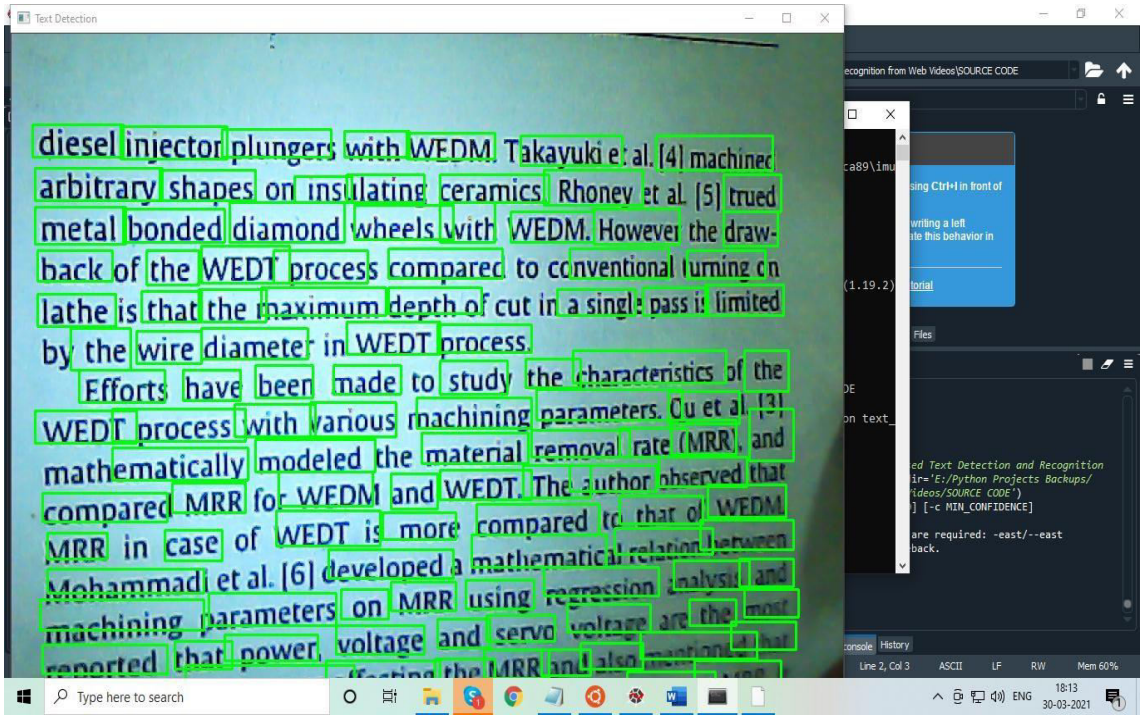


Fig 6.2: Launch The Spyder Platform

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907172 |

6.1.3 OPEN IN THE CORRESPONDENT RESULT FOLDER

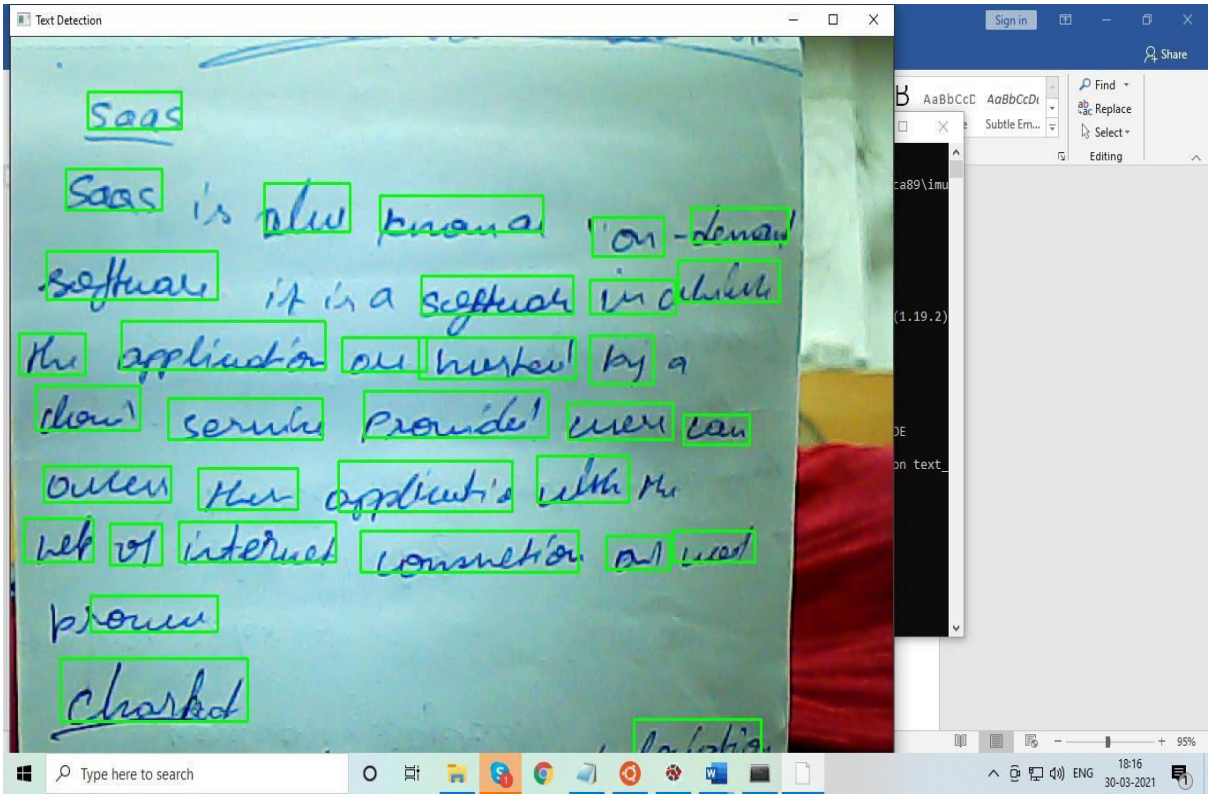


Fig 6.3 : Open In The Correspondent Result Folder

VII. CONCLUSION AND FUTURE WORK

In the past two decades, a variety of research achievements have been published in the literature. However, because of the wide variety challenges, current technologies and systems for video text detection, tracking and recognition have limited performance. Hence, typical grand challenges and recent ne applications (as described above) are ac- companies by many open issues, and numerous research opportunities for both technologies and applications. In addition to the discussions mentioned, research directions specifically for video text extraction (e.g., text tracking, and tracking based detection and recognition) are highlighted here, and possible novel applications of the technology are also discussed

REFERENCES

1. Q. Ye and D. Doermann, “Text detection and recognition in imagery: A survey,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 37, no. 7, pp. 1480–1500, 2018.
2. Y. Zhu, C. Yao, and X. Bai, “Scene text detection and recognition: Recent advances and future trends,” Frontiers in Computer Science, vol. 10, no. 1, pp. 19-36,2019.
3. Zayene, Oussama, Jean Hennebert, Sameh Masmoudi Touj, Rolf Ingold, and Najoua Essoukri Ben Amara. "A dataset for Arabic text detection, tracking and recognition in news videos-AcTiV." In *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*, pp. 996-1000. IEEE, 2015.
4. Tapu, Ruxandra, Bogdan Mocanu, and Titus Zaharia. "DEEP-SEE: Joint object detection, tracking and recognition with application to visually impaired navigational assistance." Sensors 17, no. 11 (2017): 2473.
5. Ye, Qixiang, and David Doermann. "Text detection and recognition in imagery: A survey." IEEE transactions on pattern analysis and machine intelligence 37, no. 7 (2014): 1480-1500.
6. Zhu, Yingying, Cong Yao, and Xiang Bai. "Scene text detection and recognition: Recent advances and future trends." Frontiers of Computer Science 10, no. 1 (2016): 19-36.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907172 |

7. Li, Huiping, David Doermann, and Omid Kia. "Automatic text detection and tracking in digital video." IEEE transactions on image processing 9, no. 1 (2000): 147-156.
8. Mukhametshin, Stanislav, Alisa Makhmutova, and Igor Anikin. "Sensor tag detection, tracking and recognition for AR application." In 2019 International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM), pp. 1-5. IEEE, 2019.
9. Yin, Xu-Cheng, Ze-Yu Zuo, Shu Tian, and Cheng-Lin Liu. "Text detection, tracking and recognition in video: a comprehensive survey." IEEE Transactions on Image Processing 25, no. 6 (2016): 2752-2773.
10. Xi, Jie, Xian-Sheng Hua, Xiang-Rong Chen, Liu Wenyin, and Hong-Jiang Zhang. "A video text detection and recognition system." In IEEE International Conference on Multimedia and Expo, 2001. ICME 2001., pp. 222-222. IEEE Computer Society, 2001.

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 7.542

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details