A Proposed Framework of AFS to Extract Semantic Text From Vedio Content

Mahmoud Yehia Abo El Naga
Faculty of Computers and Information, Sadat Academy for Management Sciences, Cairo, Egypt

Assoc. Prof. : Nancy Awadallah
Faculty of Computers and Information, Sadat Academy for Management Sciences, Cairo, Egypt

Abstract

In recent years, the popularity of video is increasing rapidly. This created a need of algorithms that automatically index videos-based content to facilitate the navigation through video›s libraries text in a video often gives an indication of a scene›s semantic content, Text extraction Is not an easy task due to the unconstrained nature of general – purpose videos; it may have arbitrary color, size, and orientation as well as text backgrounds may be complex and changing.

Keywords: Advanced Format Systems, Text Image Extraction, Enhancing Component, Optical Character Recognition, Joint Photographic Experts Group,

1-Introduction

Digital video is quickly gaining popularity Growing Internet bandwidth speeds are a crucial factor contributing to this tremendous rise in video., A significant human effort is needed to categorize these video data files Several approaches have been developed to facilitate these processes metadata and annotated data are needed to create referencing for facilitating the search for digital multimedia The textual information embedded in the multimedia data offers important information for multimedia data understanding therefore, it could be a good entity for keyword- based queries. Text extraction has gained a lot of attention, in the last years, in several applications. Jung et al., [Jung et al, 2004] distinguish between four main classes of text extraction applications as follows:

a) Page segmentation (which includes newspaper headlines extraction [7].

b) Address block location

c) Where to find license plate information [8]

d) Content-based image video indexing

Despite Jung›s efforts, the research in the field of designing domain independent text extraction systems is still a great challenge. The difficulties arise from the written text›s diversity in the video sources. The written text could be in different fonts, size color, orientation or alignment. Furthermore, the text could, possibly, be embedded in shaded or complex backgrounds.

2. Literature Review

Researcher Boris Epshtein, Eyal Ofek, Yonatan Wexler, Algorithms Stroke Width Calculation, Component Grouping, Recommend methods Simple, less cluttered environments, year 2010

Researcher Anastasios B. Spyropoulos, Berrin A. Yanikoglu, Algorithms FAST keypoint detector, Recommend methods Less accurate in complex scenes, year 2010

Researcher SWT with stroke width, Algorithms FAST keypoint detector, Recommend methods Sensitive to noise and complex backgrounds, year 2013

The proposed AFS Framework aims, not only to extract text from videos, but also, to recognize and preserve the text semantics in a readable format, The Framework consists of two main stages the Text Image Extraction (TIE) and Enhancing Components (EC).

TIE process is composed of several tasks including text detection; text localization; text tracking; text segmentation; binarization and character recognition.[10]

EC are the main contribution processes composed of semantic validation and past processing which are the main contributions of this research.[9]

3. AFS Framework

This research proposes a new framework named AFS that integrates text image extraction information (TIE) approach with Wordnet Library and Wordnet domain generate a semantic meta-data about the video content. A Text Extraction phase is divided into five stages as follows:

(i) Text detection seeks to locate text-containing areas inside images.[11]

(ii) The goal of text localization is to combine text areas and pinpoint precise text locations

(iii) Text tracking follows the customized text across a video's frames.

(iv) Text segmentation includes the separation of the localizes text from the images background

(v) OCR stage converts the binarized image to ACII text

Semantic phase is added to the proposed framework to achieve three processes as follows

(i) Validation: Some of the resulted words - from the OCR stage may be vague, altered cluttered or missing some characters. Therefore, the wordnet database is used to automatically verify the OCR engine›s output.[3]

(ii) Generate Meta-data: Attaches meta-data file to video file using Adobe Creative Suite.

(iii) Attach meta_data Attach meta_data file to video file using Adobe Creative Suite[1]

4. Proposed Framework of AFS

The proposed AFS was utilized creating semantic from video text The framework is shown in Figure.1.

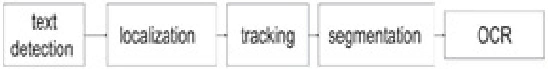


Figure.1 AFS Framework

4.1. Detect Text

This stage is responsible for deciding If an image contain text or not which were output from preprocessing task, In the computer vision [2] job of text detection, a model is trained to identify areas of a picture that could contain text. Convolutional neural networks are frequently used for this (CNNs),[5] [which learn to detect patterns and features indicative of text within the input image

4.2. Localization of Text

Text Localization is responsible for grouping text regions into text instances and generating a set of tight bounding boxes around all text instances. It applies region-based methods for text localization. Region based methods relies on the properties of the color or gray scale in a text region their differences.

4.3. Tracking

In this step will tracking the text coordinates of all images with text folder (called track regions) this result help in next step.

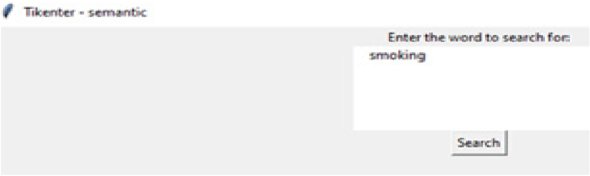
Example if have image contain text (you want) the result of these stage is list [coordinates 'you', coordinates 'want']

4.4. Text Segmentation

The goal of this stage is to place the texts extracted from the original image in the video and place them inside a new image with a white background, considering the difference between one word and another, and specifying these words inside the red box for clarification.[6]

4.5. OCR

One of the basics of this stage is searching for words in a semantic way using Wordnet databases, which contain many word vocabulary, which helps us find the video better because it is not a requirement that the word to be searched for be within the text, but rather it can be in another form that has a similar meaning.[4]



Enter the word to search for:

smoking

Figure.2 enter research word

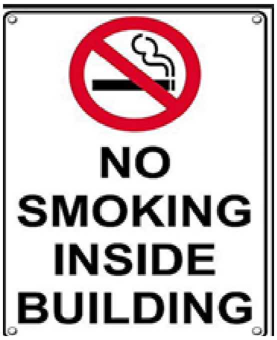


Figure.3 frame contain research word

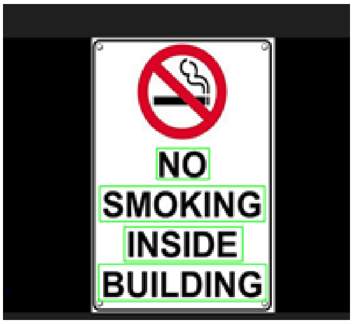
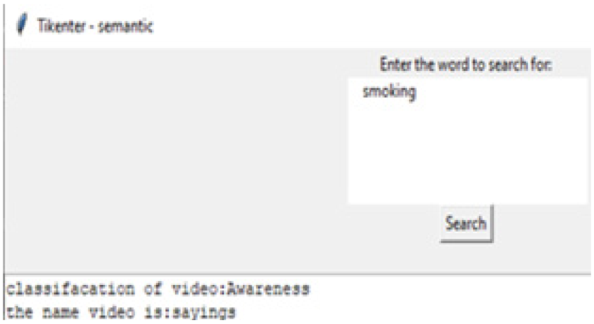


Figure.4 localization output frame



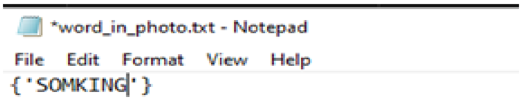
Enter the word to search for:

smoking

```
classifacation of video:Awareness
the name video is:sayings
```

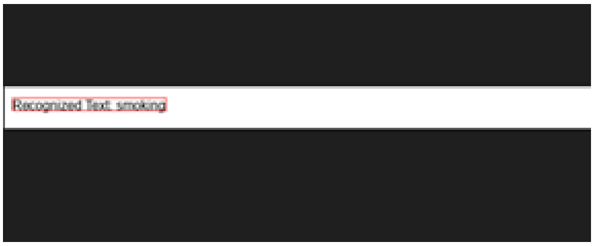
Figure.5 classification of video

5. Implementation



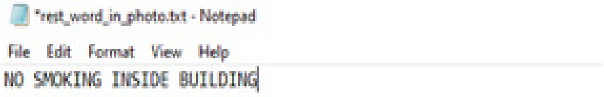
```
{'SOMKING'}
```

Figure.6 descriptive metadata research word



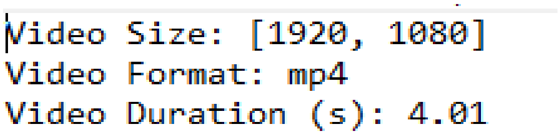
Recognized Text: smoking

Figure.7 segmentation



```
NO SMOKING INSIDE BUILDING
```

Figure.8 descriptive metadata of frame



```
Video Size: [1920, 1080]
Video Format: mp4
Video Duration (s): 4.01
```

Figure.9 technical metadata of video

5. Conclusions

This research is based on studying video search in a semantic manner, through the AFS mentioned above. There are many processes performed on the videos beforehand to improve the search process. Therefore, care must be taken in choosing these methods because they represent a noticeable change in the accuracy of the research.

.

6. References

[1]Jurafsky, Daniel, and James H. Martin. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. 3rd ed., Prentice Hall, 2021

[2] Rosebrock, A. (2021). Deep Learning for Computer Vision with Python. PyImageSearch

[3]Raj, B. (2020). Advanced Deep Learning with Keras: Apply deep learning techniques, autoencoders, GANs, variational autoencoders, deep reinforcement learning, policy gradients, and more. Packt Publishing

[4]Geron, A. (2019). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems (2nd ed.). O›Reilly Media.

[5]Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. Deep Learning. MIT Press, 2016

[6]Bradski, G., & Kaehler, A. (2008). Learning OpenCV: Computer vision with the OpenCV library. O›Reilly Media

[7] [P.K. Loo et al, 2004] P.K. Loo and C.L. Tan. Adaptive Region Growing
Color Segmentation for Text Using Irregular Pyramid. International Workshop on Document Analysis Systems, pages 264-275.Springer Verlag, 2004

[8] [Jung et al,2004] K. Jung, K.I. Kim, and A.K. Jain. «Text In form action Extract ion in Images and Videos: A Survey». Pattern Recognition Letters 31977-997, 2004

[9] Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. Journal of Machine Learning Research, 3, 993-1022.

[10] Conference on Image Analysis and Processing, Pages 192-197. IEEE Computer Society, 2001.

[11] [JF. Canny,1986] J. F. Canny. «A computational approach to edge detection». IEEE Trans. on Pattern Analysis and Machine Intelligence, :679-698, 1986.