EFFICIENT MODEL FOR PRECISE TEXT EXTRACTION FROM VIDEO USING TEXT LABELLING & DEEP LEARNING MODELS

TARUN TATAKUNTALA | B1231849 | Research Methods CIS4011-N-FJ1-2021

Research Purpose

1. Although many text detection methods are proposed and implemented for indexing and annotations, there are many other applications like supermarket automation, merchandise movement, licence plate recognition etc., These are useful for the society in terms of efficiency and speed in processing the things.
2. Text Detection gives raise to many applications, but video based text detection is a challenging task where it requires following stages: text tracking, detection, localization, and recognition. In the literature, many methodologies have been proposed in terms of text localisation, which targets to determine the position of the transcript. The process of detection is utilized to decide whether any text is present or not.
3. Text recognition is used to classify the transcript from non-transcript. Enhancement process can be encompassed to improve the resolution prior to recognition. However, applying fusion techniques to improve the resolution prior to recognition will lead to high efficiency and accuracy.
4. Previous work on text detection in video using fusion techniques had been proposed, but still need improvement which increases the recall and precision.Proposed an innovative approach which combines Laplace operator by high frequency high band wavelet across multi-level fusion to recognize candidate transcript.
5. Once images are fused, they are fed to machine to classify the text in the appropriate group. Here, we can use convolution neural network for classification of text which outperforms the existing models without additional training data in the text classification task.

Fig 1. Text Detection in a video frame

Background of Research

1. The wide popularity of videos, images, documents available in the internet have led to the demand for automatic annotations, indexing, construction of images/visuals and many applications which can fed to automation.
2. In order to implement these demands, transcript is a major cause of information which requires detection, localization, tracking and recognition process.
3. Nevertheless, transcript variation owing to font-size, font-style, direction, occlusion, poor resolution makes automatic transcript extraction more challenging.
4. We have explored the survey with different feature descriptors, methods for detection and recognition, datasets, applications and evaluation process from diverse publications.

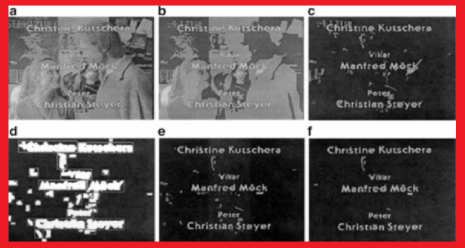

Fig 2. (a)Initial visual frame (b) Dissection phase
(c) Dimension restraint phase (d) Binary process & dilation
(e) Movement analysis (f) Result of dissimilarity analysis [1]

Research Methodology

Qualitative basis: The research was carried out in Qualitative basis in which depend on the data and information gathered from review of journal articles

Philosophical basis: This research will be developed on the basis of existing data, studies and hypothesis taken from the journals.

Data Collection: Data sets are collected from MSRA, ICDAR 2011 and ICDAR 2013 datasets which includes graphics transcripts in visual , web-based imageries and emails[2]

Data Analysis Technique: Text recognition is a process of identifying a text from scanned documents or images or video into a form which can be analysed and manipulated by a computer.

OCR (Optical character Recognition) is the costly and speedy method for identifying the characters.

Using OCR to verify the text after classifying text and non-text using a graphical model.[3]

Recent trends have transformed OCR to machine learning approach to identify the text from video. Machine learning allows creating architecture with multilayer perceptron to train the machine to categorize text and non-text.

Once they are classified, machine learning techniques can be utilized for verification of transcript from non-transcript.

The ICDAR 2013/15 datasets are prepared for text localization and end-to-end tasks. End-to-end tasks include localization of text followed by recognition of words. Following table summarizes the key modifications that were introduced for the dataset 2015. It shapes the achievement of earlier editions.

Tasks	Challenge 1: Born-Digital	Challenge 2: Focused Scene Text	Challenge 3: Text in Videos	Challenge 4: Incidental Scene Text
1. Localization / Tracking	2011 / 2013	2011 / 2013	2013 / 2015	2015
2. Segmentation	2011 / 2013	2013		
3. Recognition	2011 / 2013	2011 / 2013		2015
4. End-to-end	2015	2015	2015	2015

Evaluation Methodology

1. Detection rate is a measure of performance for natural scene extraction. It is calculated as the number of detected transcripts divided by the number of transcripts for that frame.

2. ICDAR 2003 [4] shows precision/recall measures for retrieval system. Precision, p' is termed as ratio between the number of true estimates (TE) divided by the total number of estimates (E),

$$p' = TE/E.$$

3. System which shows over-estimations of the number of transcript rectangles will acquire low precision. Recall, r' is termed as the number of true estimates (TE) divided by the total number of targets (T),

$$r' = TE/T.$$

4. System which under-estimates the number of transcript rectangles will have a low recall.

precision = number of true estimates / total number of estimates

recall= number of true estimates / total number of targets

Schedule and Publishing

The scheduled time to complete the research is 13 weeks. In order to successfully evaluate the research, the scheduling of the tasks is very important. The research of efficient model for precise text extraction from video using text labelling and deep learning models also requires the proper schedule and publishing. Below is the schedule for the research carried out.

Week 1-2:
Research's Introduction and Formation of Research Question

Week 3-5:
Research Methodology, Collection of data, Literature review.

Week 6-8:
Inclusion and Exclusion Criteria

Week 9-10:
Professional, legal and ethical issues.

Week 11-13:
Conclusion, Final draft and Submission

Professional, Legal and Ethical issues

There are abundant challenges for detection and recognition of transcript in video. Moving text, text degradation, background complexity, text variations has made video text detection to evolve. Following are the issues faced while deriving text from video.

1. Poor resolution
2. Colour bleeding
3. Moving text
4. Real time system

Apart from this, there are certain ethical issues. They are bias in the data, inaccuracy of results, legal consent violations and ethical consent violations.

References

[1] Lienhart RW, Stuber F, "Automatic text recognition in digital videos," Proc SPIE, 2666(3), pp. 180-188, 1996.

[2] Q. Y and D. Doermann, "Text detection and recognition in imagery: A survey," IEEE Trans. Pattern Anal. Mach. Intell., vol. 37, no. 7, pp. 1480–1500, Jul. 2015.

[3] A.Thilagavathy, K.Aarthi, A.Chilambuchelvan, "Text Detection and Extraction from Videos using ANN based network," in IJSCAI, Vol. 1, No. 2, Aug-2012.

[4] R. minetto, N. Thome, M.Cord, J. Fabrizio, B. Marcotegui," A multiresolution system for text detection in complex detection in complex visual scenes," IEEE Conf, vol. 1, pp 3862-3864, 2010.

[5] Y. Zhu, C. Y, and X. Bai, "Scene text detection and recognition: Recent advances and future trends," Frontiers Comput. Sci., vol. 10, no. 1, pp. 19–36, Feb. 2016.

[6] Xu-Cheng Yin, Ze-Yu Zuo, Shu Tian, and Cheng-Lin Liu, "Text Detection, Tracking and Recognition in Video: A Comprehensive Survey," IEEE Trans.Image Processing, vol.25, no. 6, June.2016.

[7] Liang Wu, Palaiahnakote Shivakumara, Tong Lu, and Chew Lim Tan, "A New Technique for Multi-Oriented Scene Text Line Detection and Tracking in Video," IEEE Trans.Multimedia, vol.17, no. 8, Aug-2015