International Journal on Future Revolution in Computer Science & Communication Engineering
Volume: 4 Issue: 11
ISSN: 2454-4248
42 – 45

Efficient Secure For Tracking Based Text Detection and Recognition from Web Videos

S. Karthi[1]
M.Phil - Research Scholar,
Department of Computer Science,
Selvamm Arts and Science College
(Autonomous),Namakkal

Mrs. D. Ananthanayaki
M.C.A., M.Phil., [2]
Assistant Professor, Department of
Computer Science,
Selvamm Arts and Science College
(Autonomous), Namakkal

Mrs. K. K. Kavitha
M.C.A., M.Phil., SET., (Ph.D).,[3]
Vice Principal, Head of the
Department of Computer Science,
Selvamm Arts and Science College
(Autonomous)
Namakkal

ABSTRACT: Sensor nodes forming a network and using wireless communications are highly useful in a variety of applications including battle field (military) surveillance, building security, medical and health services, environmental monitoring in harsh conditions, for scientific investigations on other planets, etc.

But these wireless sensors are resource constricted: limited power supply, bandwidth for communication, processing speed, and memory space. One possible way of achieve maximum utilization of those constrained resource is applying signal processing and compressing the sensor readings.

Usually, processing data consumes much less power than transmitting data in wireless medium, so it is effective to apply data compression by trading computation for communication before transmitting data for reducing total power consumption by a sensor node. However the existing state of the art compression algorithms are not suitable for wireless sensor nodes due to their limited resource.Therefore there is a need to design signal processing (compression) algorithms considering the resource constraint of wireless sensors. In our work, we designed a lightweight codec system aiming surveillance as a target application.

In designing the codec system, we have proposed new design ideas and also tweak the existing encoding algorithms to fit the target application. Also during data transmission among sensors and between sensors and base station, the data has to be secured. We have addressed some security issues by assessing the security of wavelet tree shuffling as the only security mechanism.

*****

I. INTRODUCTION

A means to naturally recognizing and fetching out the content of video description would possibly make them indexed in considerable and appropriate way for later reference, and would facilitate actions viz. automatic notification and dissemination, to be triggered in real time by the contents of streaming video. Video text recognition, or video OCR, is a constructive tool to characterize the contents of video containing overlay text (text captions superimposed over the video imagery, such as in broadcast news programs) and scene text (text that appears in the real scene of the video, such as text on street signs, nameplates, and billboards). In this paper exhaustive survey is done for text detecting, extraction and recognizing in complex images and video frames.

Digital Videos are widely used both professionally and domestically because of the easy availability of camcorders to mobile phones. People are increasingly making videos may be for commercial use or personal use, this is leading to growing content of Video. While we can capture, compress, store, transmit and display video with great facility, editing videos and manipulating them based on their content is still a non-trivial activity. This paper concentrates on extracting text out of the video frames, taken out of the video.This property has made text detection and recognition in natural images active research topics in computer vision. Since scene text images usually contain a large amount of irrelevant elements in addition to useful text contents, text detection is a critical procedure to localize texts and discard irrelevant elements. Then text recognition is required to interpret the symbols in the localized text regions and to convert them into computer readable and tasks or treat them as two isolated stages in an end-to-end text recognition system. In this paper, we propose a unified framework, which takes text detection and recognition as a whole and performs both tasks in a single unified pipeline.

The explosive growth of smart phones and the online social media have led to the accumulation of large amounts of visual data, in particular, the massive and increasing collections of video on the Internet and social networks. These countless web videos have triggered research activities in multimedia understanding and video retrieval, where text in video contains valuable information and is exploited in widespread content-based video applications

 ISSN: 2454-4248
42 – 45

II. LITERATURE SURVEY

Title: Texture-based approach for text detection in images using support vector machines and continuouslyadaptive mean shift algorithm
Author:K. I. Kim, K. Jung, and J. H. Kim,
Published in:IEEE Transactions on Pattern Analysis and Machine Intelligence (Volume: 25, Issue: 12, Dec. 2003)

The current paper presents a novel texture-based method for detecting texts in images. A support vector machine (SVM) is used to analyze the textural properties of texts. No external texture feature extraction module is used, but rather the intensities of the raw pixels that make up the textural pattern are fed directly to the SVM, which works well even in high-dimensional spaces. Next, text regions are identified by applying a continuously adaptive mean shift algorithm (CAMSHIFT) to the results of the texture analysis. The combination of CAMSHIFT and SVMs produces both robust and efficient text detection, as time-consuming texture analyses for less relevant pixels are restricted, leaving only a small part of the input image to be texture-analyzed.

Title:A comprehensive method for multilingual video text detection, localization, and extraction
Author:M. R. Lyu, J. Song, and M. Cai,
Published in:IEEE Transactions on Circuits and Systems for Video Technology (Volume: 15, Issue: 2, Feb. 2005)

Text in video is a very compact and accurate clue for video indexing and summarization. Most video text detection and extraction methods hold assumptions on text color, background contrast, and font style. Moreover, few methods can handle multilingual text well since different languages may have quite different appearances. This paper performs a detailed analysis of multilingual text characteristics, including English and Chinese. Based on the analysis, we propose a comprehensive, efficient video text detection, localization, and extraction method, which emphasizes the multilingual capability over the whole processing.

The proposed method is also robust to various background complexities and text appearances. The text detection is carried out by edge detection, local thresholding, and hysteresis edge recovery. The coarse-to-fine localization scheme is then performed to identify text regions accurately. The text extraction consists of adaptive thresholding, dam point labeling, and inward filling. Experimental results on a large number of video images and comparisons with other methods are reported in detail.

III. RESEARCH METHODOLOGY

EXISTING SYSTEM

The basic level of multimedia retrieval on the basis of textual information is keyword searching. Numerous methods have been proposed in solving documents indexing and retrieval tasks based on only text content with noisy data. These methods can improve the retrieval performance on top of simple word matching using fuzzy logic, confusion information for characters and a bi-gram model, finite state machine, similarity distance measure, and OCR error modeling.In the existing the static image can be used to extract the text. In that a static is to retrieval of text from the image. It is not in the order format. So, it is difficult to read the what is text in it.

DISADVANTAGES

- The accuracy of the text is low from the live and web videos.
- It take more time

IV. PROPOSED SYSTEM

In a wide variety of approaches have been proposed for text detection and recognition in video. On the one hand, the most direct and simple way is to recognize video text as the same as the one in static images, i.e., to recognize text with frame by frame. Hence, conventional video text extraction techniques mainly focus on detecting and recognizing text in each individual frame or some key frames, without multi-frame integration. On the other hand, spatial and temporal information is very important for multimedia understanding of complex videos. Consequently, there are also some video text detection and recognition methods with tracking techniques using multiple frames. Most existing tracking based text detection and recognition methods can be roughly categorized into temporal-spatial based methods and fusion based ones. The former methods use temporal or spatial information to remove noises for detection or to enhance the images for recognition. The latter ones merge detection and tracking results, or recognition and tracking results over multiple frames. However, the feedback between tracking and detection or recognition is always ignored. How to effectively utilize the relations and interactions between tracking and detection or recognition is challenging for text extraction from complex videos.

ADVANTAGES

- Text tracking will advance a great deal as a key component for text extraction from complex videos in the future.
- We can get accurate text from web videos.

STATEMENT OF PROBLEM

The explosive growth of smart phones and the online social media have led to the accumulation of large amounts of visual data, in particular, the massive and increasing collections of video on the Internet and social networks. These countless web videos have triggered

research activities in multimedia understanding and video retrieval, where text in video contains valuable information and is exploited in widespread content-based video applications. In a wide variety of approaches have been proposed for text detection and recognition in video.

On the one hand, the most direct and simple way is to recognize video text as the same as the one in static images, i.e., to recognize text with frame by frame. Hence, conventional video text extraction techniques mainly focus on detecting and recognizing text in each individual frame or some key frames, without multi-frame integration. On the other hand, spatial and temporal information is very important for multimedia understanding of complex videos. Consequently, there are also some video text detection and recognition methods with tracking techniques using multiple frames. Most existing tracking based text detection and recognition methods can be roughly categorized into temporal-spatial based methods and fusion based ones. The former methods use temporal or spatial information to remove noises for detection or to enhance the images for recognition. The latter ones merge detection and tracking results, or recognition and tracking results over multiple frames. However, the feedback between tracking and detection or recognition is always ignored. How to effectively utilize the relations and interactions between tracking and detection or recognition is challenging for text extraction from complex videos.

HARDWARE REQUIREMENTS

The hardware requirements may serve as the basis for a contract for the implementation of the system and should therefore be a complete and consistent specification of the whole system. They are used by software engineers as the starting point for the system design. It should what the system do and not how it should be implemented.

HARDWARE REQUIREMENT

- Processor :- Pentium Dual Core 2.00GHZ
- Hard disk :- 40 GB
- Mouse :-Logitech.
- RAM :- 2GB(minimum)
- Keyboard : -110 keys enhanced.

SOFTWARE REQUIREMENTS

The software requirements document is the specification of the system. It should include both a definition and a specification of requirements. It is a set of what the system should do rather than how it should do it.

The software requirements provide a basis for creating the software requirements specification. It is useful in estimating cost, planning team activities, performing tasks and tracking the teams and tracking the team's progress throughout the development activity.

SOFTWARE REQUIREMENT

- Operating system :- Windows7 , 8.1
- Front End :- Microsoft Visual Studio .Net 2013
- Coding Language :- C#
- Backend :- SQL Server 2012

MODULES

1. VIDEO ACQUISITION OR BROWSING
2. PREPROCESSING
3. EXTRACT FRAMES FROM VIDEOS
4. EXTRACT TEXT FROM FRAMES
5. LOG THE EXTRACTED TEXT

MODULES DESCRIPTION

VIDEO ACQUISITION OR BROWSING

In this module, user get image from web camera or mobile phone images or captured image. Based on this image, users get text for further processing. And input image may be in RGB format or grey scale image. We can import the video from the users and convert video into frames. Each frame is considered as image. Video is automatically converted into frames after every 5 seconds.

PREPROCESSING

In this module, RGB image is converted into grayscale image and eliminate noises from images usingmedian filter. Finally apply image Binarizationtechnique to separate background and foregroundimage.

EXTRACT FRAMES FROM VIDEOS

VideoCEL is basically a library for video content extraction. Its components extract relevant features of video data and can be reused by different applications. The object model includes components for video data modelling and tools for processing and extracting video content, but currently the video processing is restricted to images.

At the data modelling level, the more significant concepts are the following:

- Images, for representing the frame data, a numerical matrix whose values can be colors, colormap entries, etc.;
- ColorMaps, which map entries into a color space, allowing an additional indexation level;
- ImageDisplayConvertes and ImageIOHandlers, that convert images in the specific formats of the platforms and vice-versa.

 ISSN: 2454-4248
42 – 45

Tools for data processing are applied to the described data modelling classes, and also modelled as a hierarchy of classes: the ImageOPs, These operators represent functions which are applied to image regions and extract "single-image" or sequential content features. The implemented algorithms and procedures are described in more details in the next sections.

EXTRACT TEXT FROM THE FRAME

The problem of Text Information Extraction needs to be defined more precisely before proceeding further. A TIE system receives an input in the form of a still image or a sequence of images. The images can be in gray scale or color, compressed or un-compressed, and the text in the images may or may not move.

The TIE problem can be divided into the following subproblems:

(i) Detection
(ii) Localization
(iii) Tracking,
(iv) Extraction and enhancement
(v) Recognition

V. CONCLUSION

In this paper, we propose a generic Bayesian-based framework of Tracking based Text Detection And Recognition from web videos for embedded captions. This framework includes three major components, i.e., text tracking, tracking based text detection, and tracking based text recognition. For tracking based text detection, a tracking by- detection method is used to track the text and a revising mechanism is designed to improve the recall and precision of text detection.

In general, temporal redundancy in video is helpful to improve the performances of text detection and recognition. Text tracking will advance a great deal as a key component for text extraction from complex videos in the future. However, object tracking is an opening issue. In particular, ID switch always occurs inevitably in complex situations. Hence, text tracking in complex videos (e.g., web videos) is still a challenging topic, and post-processing techniques of text tracking should be further investigated.

FUTURE ENHANCEMENT

In future we implement this concept for share and download the text converts text into voice, and send into other user with short span of time. The voice may be recorded, downloaded or share it to others using player or speaker

REFERENCE

[1] images and video:a survey," Pattern Recognition, vol. 37, no. 5, pp.977–997, 2004. 1, 2, 3

[2] J. Liang, D. Doermann, and H. Li, "Camera-based analysis of text and documents: a survey," International Journal of DocumentAnalysis and Recognition, vol. 7, no. 2-3, pp. 84–104, 2005.

[3] 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, 2015.

[4] X.-C. Yin, Z.-Y.Zuo, S. Tian, and C.-L. Liu, "Text detection, tracking and recognition in video: A comprehensive survey," IEEETransactions on Image Processing, vol. 25, no. 6, pp. 2752–2773,2016.

[5] B. Wang, C. Liu, and X. Ding, "A research on video text tracking and recognition," in Proceedings of SPIE, 2013, vol. 8664.