

Semantic Multimedia Annotation: Text Analysis

Irfanullah^{*1,2}, Nida Aslam^{*1,2}, Kok-Keong Loo^{*1}, Roohullah²

^{*1, Corresponding author} School of Engineering and Design, Brunel University, United Kingdom

² Kohat University of Science and Technology (KUST), Kohat, N.W.F.P, Pakistan

(Irfan.Khan, Nida.Aslam, Jonathan.Loo)@brunel.ac.uk, roohullah_orc@yahoo.com

doi: 10.4156/jdcta.vol3.issue2.irfanullah

Abstract

The demand for fast access of multimedia content is growing exponentially, but unfortunately the advancement in technology is not so fast. A variety of annotation techniques are proposed by the researcher in literature, but still the demand for general annotation mechanism is exists. Among different techniques available for annotation the text analysis is the first hand technique that allows the machine to produce direct high-level semantics with computation. In this paper, we have discussed the existing state-of-the-art techniques available for the text analysis in image/video.

Keyword

Video/Image OCR, Text Analysis

1. Introduction

From the last decade the multimedia information usage and production are growing day by day. In all media types video is the most informative and challenging as it is a combination of all other media. The way by which the video is presented for access has become a challenging task both for application system and the viewers. So the need for the extraction of metadata from the video is high which help the system/tools in retrieval the specific video more efficiently.

A vast variety of techniques are proposed in literature for video analysis ranging from low-level features to high-level semantic extraction and all these techniques are based on colour, texture, shape, sound, text, objects and the like techniques.

The annotation based on text in video gaining the interest of the researchers due to its nature, as text is almost the direct way to explain the video content and more helpful in preparing the metadata related to video. The text is almost embedded in every type of videos

and provides useful information about the video or video segment. For commercial production the text is a mandatory part of the videos as it describes all the contents and provides a platform for the viewer to understand easily the video/video segment.

The text extracted by the Video/image OCR does not directly use for the analysis of the annotation process related to the video/image, as it does not provide a comprehensive information about the video. But it cannot be ignore for the annotation as it contains useful information about the video either directly or indirectly as the text is very useful for describing the content of an image or videos, and the successful extraction of text enables automatic text based annotation and subsequent keyword-based searching or content-oriented process.

Off all the available techniques for the video annotation, only text analysis is useful for the high-level semantic directly, while other techniques require an extra effort to produce high-level semantic.

2. Types of Text in Video/Image

The text in video and image can broadly be categorized in two main groups according to the text appear in the video/image.

2.1. Artificial text: The text that is superimposed on the image/video frame at the time of editing. Usually this type of text is laid over the video/image at the later stage as in figure 1. It mostly includes video title, production company and information about the videos which are directly useful for metadata.

2.2. Scene text: A text that is appear in the video frame but not added during editing while it a part of the scene during video capturing. The product name, Banners, advertisement about products appears in the video are the example of scene text. The figure 2 shows the scene text.



Figure 1. Artificial Text



Figure 2. Scene Text

3. Characteristics of Text

Text in images can exhibit many variations with respect to the following properties [1].

Table 2: Text Characteristics

Property	Variant or Sub-classes
Geometry	Size <ul style="list-style-type: none"> • Regularity in size of text
	Alignment <ul style="list-style-type: none"> • Horizontal / Vertical • Straight line with skew (implies vertical direction) • Curves • 3D perspective distortion
	Inter-character distance <ul style="list-style-type: none"> • Aggregation of characters with uniform distances
	Aspect-Ratio
	Strokes <ul style="list-style-type: none"> • Different stroke density and statistics
Color relationship	Color <ul style="list-style-type: none"> • Gray • Color (monochrome, polychrome)
	Text Polarity <ul style="list-style-type: none"> • Text color is dark or light

Motion	<ul style="list-style-type: none"> • Static • Linear movement • 2D rigid constrained movement • 3D rigid constrained movement • Free movement
Edge	<ul style="list-style-type: none"> • Strong contrast (edges) at text boundaries
Compression	<ul style="list-style-type: none"> • Un-compressed image • JPEG, MPEG-compressed image
Background	<ul style="list-style-type: none"> • The background can be different

4. Text Analysis Steps

The propose method for the text analysis in videos/image is proposed by many researchers as shown in Figure 3.

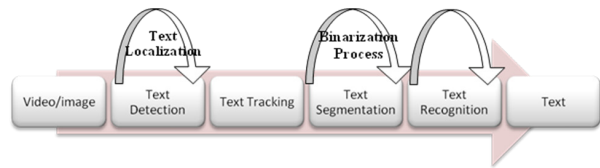


Figure 3. Four steps of the Text Analysis

It's a sequential modular approach in which a stream of video/ image is taken as input and a number of processes are perform on it with the aim to produce text from the entire video frames/image.

4.1. Text Detection

At first the input video/image are proceed for the text detection in image and/or video frame, where a recursive process text localization is perform to detect the location of the text in an image/video frame and generating the bounding box around it. Many researchers have proposed the “refinement” after localization. Text detection can be classified into two methods.

4.1.1. Bottom-up method

These methods segment the image into regions and group “character” region into words, i.e. from character → word

4.1.2. Top-Down Method

In this, 1st detect text regions are performed and then by using bottom-up technique inside the region. Top-down approach again is divided into two approaches .i.e. Heuristic method, Machine learning methods

4.2. Text Tracking

It is useful for text enhancement and detection verification. In video where same text is spread over multi-frames. According to [4] text is either stationary or linearly moving and even stationary text may move by some pixels around its original position from frame to frame. [5, 6] utilize motion vectors in MPEG-I bit stream in a compressed domain for text tracking. The effort of [7] for text tracking plays a vital role as he describes each and every factor for text tracking in consecutive frames.

4.3. Text Segmentation

Poor quality text can be difficult to extract especially with complex background and low contrast. Text segmentation includes quality enhancement and image binarization. [8] uses a linear interpretation technique to magnify small text at a higher resolution for commercial OCR software. [9, 10] present a different approach for text enhancement in the selected segment. While [11] used a multi-frame text enhancement technique by using mean square errors of two consecutive text blocks and motion train information. The [12] adaptive thresholding method is used to filter out non-text regions from text segment. [13] proposed a text enhancement method which uses a multi-hypothesis approach.

4.4. Text Recognition

Researcher used commercial OCR machines to recognize the detected and segmented text images. [14] uses Recognita OCR, [15] use a conventional pattern matching technique to recognize character.

5. Techniques for text detection in video/image

Mostly the following techniques are used for text detection.

5.1. Region based

This method employs the Connected Component analysis, which is based on the analysis of the geometrical arrangement that belong to character. The region based text detection is divided into two classes

5.1.1.Connected-Component Based (CC)

This method segments the image into regions and group “character” region into words (bottom-up technique). [16] the basic elements are created using the similarity of neighbor pixels in color, size and apply motion analysis to enhance text extraction results. While [17] uses region growing method by giving a seed region.

5.1.2.Edge-Based

[18] focus on the high contrast between the text and the background, [19] uses edge-based text detection using morphological procedures and efficiently detect horizontal artificial texts, while [20] uses edge-based text detection followed by a conditional dilation technique to choose text and inverted text objects.

5.2. Texture based

Every text present in image exhibits some textual properties, which may be used to distinguish it from the background. Gabor filters, Wavelets, Fast Fourier transformation, etc. are usually used to extract the textural properties of a text region in an image. [21] perform TD on wavelet-based feature extraction and neural network for texture analysis, while [22] use of an SVM trained on wavelets features. The [23] uses the DCT coefficients of compressed jpeg and mpeg files for distinguishing between the texture of textual and not-textual regions. [24] perform TD in video frames by segmenting the image using color image edge detector and then classifying the text/not-text blocks with the help of artificial neural network using features obtained by Gabor filtering.

5.3. Other techniques

The techniques other than CC and edge-based are, [25] propose a learning based approach, based on two stages text region extraction and text verification, where [26] propose a methodology in which the text regions are located, then by using spatial filters for removing noisy regions.

6. Open Issues of Text Analysis

Although a lot of work has been done by the researcher, but it is still not easy to design domain independent text extraction system, this is because, there are so many possible sources of variation when extracting text with different font, size, color, orientation and alignment, which could possibly be embedded in shaded or complex background.

1. Extraction of Scene text
2. Artificial Text detection with special effect (angle, color changes, size changing, new style, fonts)
3. Artificial text written in different language
4. Text tracking for non-linear moment of text like (text rotation, zoom in/out)
5. Binarization in complex background

7. Conclusions

Text analysis in video/image provides an opportunity to extract high-level semantics from the video/image with less computational effort as compared with other techniques. But due to the advancement in graphics and technology, a variety of text presentation styles are used in image/video which lay down a stone hurdles in the progress of text analysis in video/image.

8. References

1. K. Jung, K.I. Kim, and A.K. Jain. Text Information Extraction in Images and Videos: A Survey. *Pattern Recognition Letters*, 37:977–997, 2004.
2. Rainer Lienhart and Frank Stuber, (1995) «Automatic text recognition in digital videos», Technical Report / Department for Mathematics and Computer Science, University of Mannheim ; TR-1995-036
3. H. Li and D. Doermann. Text Enhancement In Digital Video Using Multiple Frame Integration. *Proceedings of ACM Multimedia 99* , pages 19-22
4. Rainer Lienhart and Frank Stuber, (1995) «Automatic text recognition in digital videos», Technical Report / Department for Mathematics and Computer Science, University of Mannheim ; TR-1995-036
5. S. Antani, U. Gargi, D. Crandall, T. Gandhi, and R. Kasturi, “Extraction of Text in Video”, Technical Report of Department of Computer Science and Engineering, Penn. State University, CSE-99-016, August 30, 1999.
6. U. Gargi, D. Crandall, S. Antani, T. Gandhi, R. Keener, and R. Kasturi, “A System for Automatic Text Detection in Video”, *Proc. of International Conference on Document Analysis and Recognition*, 1999, pp. 29 – 32.
7. H. Li and D. Doermann. Text Enhancement In Digital Video Using Multiple Frame Integration. *Proceedings of ACM Multimedia 99* , pages 19-22
8. T. Sato, T. Kanade, E. Hughes, and M. Smith , “Video OCR for Digital News Archives”, *IEEE Workshop on Content-Based Access of Image and Video Databases(CAIVD’98)*, January, 1998, pp. 52 - 60.
9. H. Li, D. Doermann, and O. Kia, “Automatic Text Detection and Tracking in Digital Video”. *IEEE Transactions on Image Processing - Special Issue on Image and Video Processing for Digital Libraries*, pages 147-155, 1999
10. H. Li and D. Doermann, “A Video Text Detection System based on Automated Training”, *Proc. of IEEE International Conference on Pattern Recognition*, 2000, pp. 223-226.
11. H. Li, O. Kia, and D. Doermann, Text Enhancement in Digital Video, *Proc. of SPIE, Document Recognition IV*, 1999, pp. 1-8.
12. [Niblack-86] W. Niblack, “An Introduction to Digital Image Processing”, PrenticeHall, Englewood Cliffs, NJ, 1986 pp. 115--116.
13. D. Chen, J. Odobez, and H. Bourlard, Text Segmentation and Recognition in Complex Background Based on Markov Random Field, *Proc. of International Conference on Pattern Recognition*, 2002, Vol. 4, pp. 227-230.
14. R. Lienhart and W. Effelsberg, “Automatic text segmentation and text recognition for video indexing,” *Multimedia Syst.*, vol. 8, pp. 69–81, Jan. 2000.
15. T. Sato, T. Kanade, E. Hughes, and M. Smith , “Video OCR for Digital News Archives”, *IEEE Workshop on Content-Based Access of Image and Video Databases(CAIVD’98)*, January, 1998, pp. 52 - 60.
16. J.C. Shim, C. Dorai, and R. Bolle. Automatic Text Extraction from Video for Content-Based Annotation and Retrieval. *Proceedings of the International Conference on Pattern Recognition*, pages 618–620. IEEE Press, 1998.
17. K. Sobottka, H. Bunke, and H. Kronenberg. Identification of text on colored book and journal covers. In *Proc. Int. Conf. on Document Analysis and Recognition*, pages 57–63, 1999.

18. D. Chen, K. Shearer, and H. Bourlard. Text Enhancement with Asymmetric Filter for Video OCR. Proceedings of the International Conference on Image Analysis and Processing, pages 192–197. IEEE Computer Society, 2001.
19. Malobabic J, O'Connor N, Murphy N, and Marlow S. “Automatic Detection and Extraction of Artificial Text in Video.” WIAMIS 2004 - 5th International Workshop on Image Analysis for Multimedia Interactive Services, Lisbon, Portugal, 21-23 April 2004.
20. [Perantonis-04] S. J. Perantonis, B. Gatos, V. Maragos, V. Karkaletsis and G. Petasis, "Text Area Identification in Web Images", Proc of the 3rd Hellenic Conference on Artificial Intelligence, Lecture Notes in Artificial Intelligence (3025), pp. 82-92, Samos, Greece, May 2004.
21. [Li-00a] Huiping Li, David Doermann, “A Closed-Loop Training System for Video Text Detection”, Cognitive and Neural Models for Word Recognition and Document Processing, World Scientific Press, 2000.
22. Qixiang Ye, Qingming Huang, Wen Gao, Debin Zhao: Fast and robust text detection in images and video frames. *Image Vision Comput.* 23(6): 565-576 (2005).
23. Yu Zhong, HongJiang Zhang, Anil K. Jain: Automatic Caption Localization in Compressed Video. *IEEE Trans. Pattern Anal. Mach. Intell.* 22(4): 385-392 (2000)
24. Y. Hao, Z. Yi, H. Zengguang, and T. Min. Automatic Text Detection in Video Frames based on Bootstrap Artificial Neural Network and CED. *International Journal of WSCG*, 11th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision, 11(1), 2003
25. Datong Chen, Kim Shearer and Herve Bourlard, «Extraction of special effects caption text events from digital video» *IJDAR*(5), No. 2-3, April 2003, pp. 138-157
26. Du, Yingzi, Chang, Chein-I Thouin, Paul D. “Automated system for text detection in individual video Images”, *Journal of Electronic Imaging*, 12(3), 410 - 422. 2003.