News Video Indexing and Retrieval using Overlay Text

Nilesh Bhojne[1], Pravinkumar Kamde[2] and Dr. S. P. Algur[3]

[1, 2] Department of Computer Engineering, Sinhgad College of Engineering, Pune,

University of Pune,Maharashtra State 411041, India

[1]nilesh.bhojne@gmail.com, [2]pravinkamde@rediffmail.com

[3]Department of Computer, RC University Belguam, Karnataka, India.

sidhu_p_algur@hotmail.com

ABSTRACT

From the last decade the Multi/Rich Media information production and usage is growing extremely. Video is the most informative and challenging as it is a combination of all other media. The way by which the video is present for access has become a challenging task both for system and the viewer applications. At all stages and for all target groups, an effective and efficient video retrieval facility is becoming necessary. In current video search, the search results are influenced by the metadata information such as title, captions associated with them. Our main focus of work is the notion of a semantic concept: an objective linguistic description of an observable entity. This can be very well achieved by labeling combinations of anchor person, persons and objects appearing in news, events appearing in the audiovisual content. News captions are most often found to contain information about the news being telecasted. In this paper, we present news video retrieval solution that target specific news videos based on their contents described by overlay text. The proposed approach is based on use of overlay text that conveys direct meaning of video as a source of complementary information. The whole process is divided in to two steps. Firstly, we build the "metadata labels" by detecting and extracting the overlay text. Secondly, these labels are then used to index the news videos. The experiments are carried on the news videos from NDTV News and large data set of video images containing artificial text developed at Image Processing Center (IPC) a research facility at - National University of Sciences and Technology (NUST), Pakistan..

Keywords- *News Video, Overlay Text, Video OCR, Transition map*

1. INTRODUCTION

Digital video libraries and archives of immense size are becoming accessible over data networks.Efficient video retrieval and browsing has become crucially important. Understanding the semantic contents of the video and using them for indexing is inevitable. Automatic indexing and retrieval of video information based on content is very challenging research area [1]. The most difficult problem is: what does content mean? Or, more specifically, how should one characterize visual or auditory content present in a video?How to extract them for building useful, high-level annotations that willenable content-based indexing and retrieval of video segments from huge digital video libraries [3]?Tremendous work is has been carried out towards developing automatic video searching system in recent years, however, because of the numerous

Natarajan Meghanathan, et al. (Eds): SIPM, FCST, ITCA, WSE, ACSIT, CS & IT 06, pp. 11–17, 2012.

DOI : 10.5121/csit.2012.2302

video program variations, it is still a very difficult work to design a general-purpose system for all types of video programs [14].

A vast variety of techniques are proposed in literature for video analysis ranging from extraction of low-level features to high-level semantic features and all these techniques are based on color, texture, shape, sound, text, and objects. Of all the available techniques of 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 semantics [6].

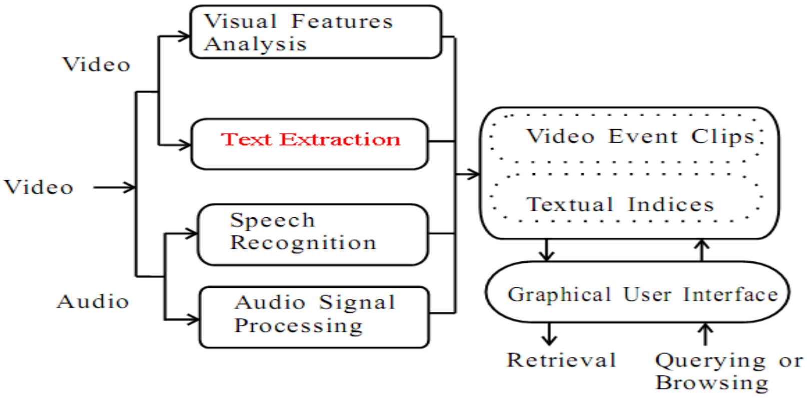


Figure 1 General Mode for Video Analysis

Text displayed in the videos can be classified into scene text and overlay text [5]. Scene text occurs naturally in the background as a part of the scene, such as the advertising boards, banners, and so on. In contrast to that, overlay text is superimposed on the video scene and used to help viewers' understanding. Since the overlay text is highly compact and structured, it can be used for video indexing and retrieval [6]. Overlay text brings important semantic clues in video content analysis such as video information retrieval and summarization, since the content of the scene or the editor's intention can be well represented by using inserted text [8]. Most broadcasting news videos tend to increase the use of overlay text that usually represent names of anchor person, place, person, or description of news in crisp. Moreover in sports news it may be name of player, type of sport, location, score and many more.

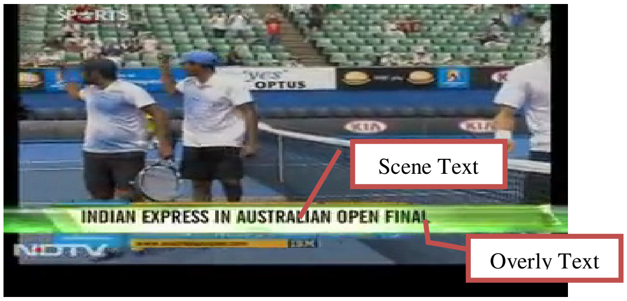


Figure 2 Types of Text in Video

2. PROPOSED METHODOLOGY

The proposed work is based on use of transition map proposed by Kim and Kim to detect the candidate text region. It is further extended to recognize the extracted text using commercial OCR provided in .NET frame work. The work is divided in to two phases. Firstly, overlay text detection is carried out using the transition region between the overlay text and background. The transition map is generated based on observation that there exist transient colors between overlay text and its adjacent background [7]. Then the overlay text regions are roughly detected by computing the density of transition pixels and the consistency of texture around the transition pixels. The detected overlay text regions are localized accurately using the projection of transition map with an improved color-based thresholding method to extract text strings correctly.Transition map method is applied to domain specific videos i. e. News Videos to extract text strings. The complementary resource i. e. overlay texts in videos are populated into an annotation database which serves for video retrieval.

The rest of this paper is organized as follows. The steps used in the process of text recognition are overlay text detection using the transition map and refine the detected text regions this is explained in Section III. The overlay text extraction from the refined text regions is explained in Section IV. The experimental results on various videos are shown in Section V, followed by conclusion in Section VI.

3. OVERLAY TEXT REGION DETECTION

The method used here is based on observations that there exist transient colors between overlay textand its adjacent background [7] as shown in Figure 3. Overlay texts have high saturation because they are inserted by using graphic components.It can also be noted that, if the background of overlay text is dark, then the overlay text tends to be bright. On the contrary, the overlay text tends to be dark if the background of overlay text is bright. Therefore, there exist transient colors between overlay text and its adjacent background due to color bleeding, the intensities at the boundary of overlay text are observed to have the logarithmical change. This also can be observed from Figure 3.

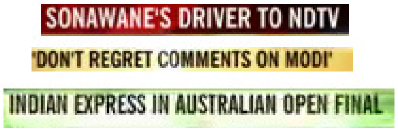

SONAWANE'S DRIVER TO NDTV
'DON'T REGRET COMMENTS ON MODI'
INDIAN EXPRESS IN AUSTRALIAN OPEN FINAL

Figure 3 Overly Text in News Videos

Since the change of intensity at the boundary of overlay text may be small in the low contrast image, to effectively determine whether a pixel is within a transition region, the modified saturation is first introduced as a weight value based on the fact that overlay text is in the form of overlay graphics. The modified saturation is defined as follows:

$$S(x,y) = 1 - \frac{3}{(R+G+B)[\min(R,G,B)]} \tag{3.1}$$

$$\tilde{S}(x,y) = \frac{S(x,y)}{\max(S(x,y)}$$

$$where\ \max(S(x,y)) = \begin{cases} 2 \times \left(0.5 \times \tilde{I}(x,y)\right), & if\ \tilde{I}(x,y) > 0.5 \\ 2 \times \tilde{I}(x,y), & otherwise \end{cases} \tag{3.2}$$

S(x, y) and max(S(x, y)) denote the saturation value and the maximum saturation value at the corresponding intensity level, respectively. $\tilde{I}(x,y)$denotes the intensity at the (x, y), which is normalized to [0, 1]. Based on the conical HSI color model, the maximum value of saturation is normalized in accordance with $\tilde{I}(x,y)$ compared to 0.5 in eq. 3.2. The transition can thus be defined by combination of the change of intensity and the modified saturation as follows:

$$D_L(x,y) = \left(1 + dS_L(x,y)\right) \times |I(x-1,y) - I(x,y)|$$

$$D_H(x,y) = \left(1 + dS_H(x,y)\right) \times |I(x,y) - \ I(x+1,y)$$

$$where\ dS_L(x,y) = \left|\tilde{S}(x-1,y) - \tilde{S}(x,y)\right| \quad and$$

$$dS_H(x,y) = \ \left|\tilde{S}(x,y) - \tilde{S}(x+1,y)\right| \qquad (3.3)$$

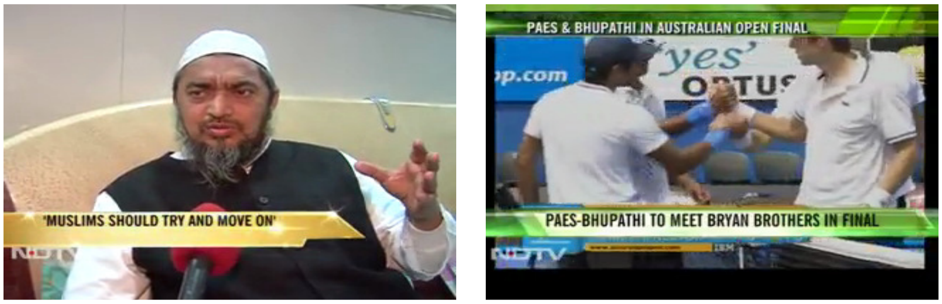
Since the weight $dS_L(x, y)$ and $dS_H(x, y)$ can be zero by the achromatic overlay text and background, 1 is added to the weight in eq. 3.3. If a pixel satisfies the logarithmical change constraint given in eq. 3.4, three consecutive pixels centered by the current pixel are detected as the transition pixels and the transition map is generated.The thresholding value TH is empirically set to 80 in consideration of the logarithmical change

$$T(x,y) = \begin{cases} 1, & if D_H > D_L + TH \\ 0, & otherwise \end{cases} \qquad (3.4)$$

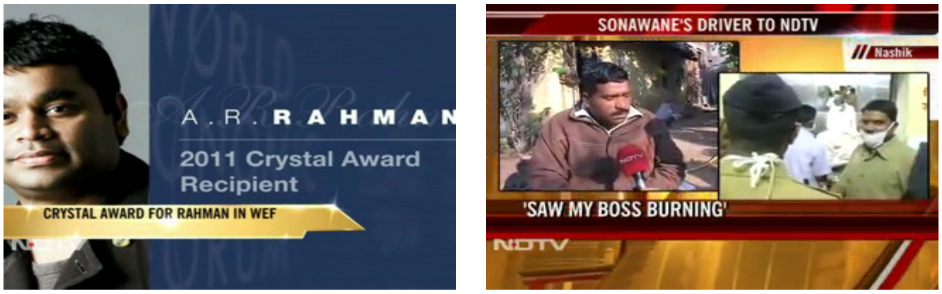
The transition map can be utilized as a useful indicator for the overlay text region. To generate the connected components, first a linked map is generated. If a gap of consecutive pixels between two nonzero points in the same row is shorter than 5% of the image width, they are filled with 1s. If the connected components are smaller than the threshold value, they are removed. The threshold value is empirically selected by observing the minimum size of overlay text region. Then each connected component is reshaped to have smooth boundaries since it is reasonable to assume that the overlay text regions are generally in rectangular shapes. As per the observations of data set used, it is noticed that the vertical height of candidate region is uniform. So the falsely detected regions with comparatively small or large heights can be eliminated. Thus the results are further refined.

4. OVERLAY TEXT EXTRACTION

Before applying video OCR application, overlay text regions need to be converted to a binary image, where all pixels belonging to overlay text are highlighted and others suppressed.

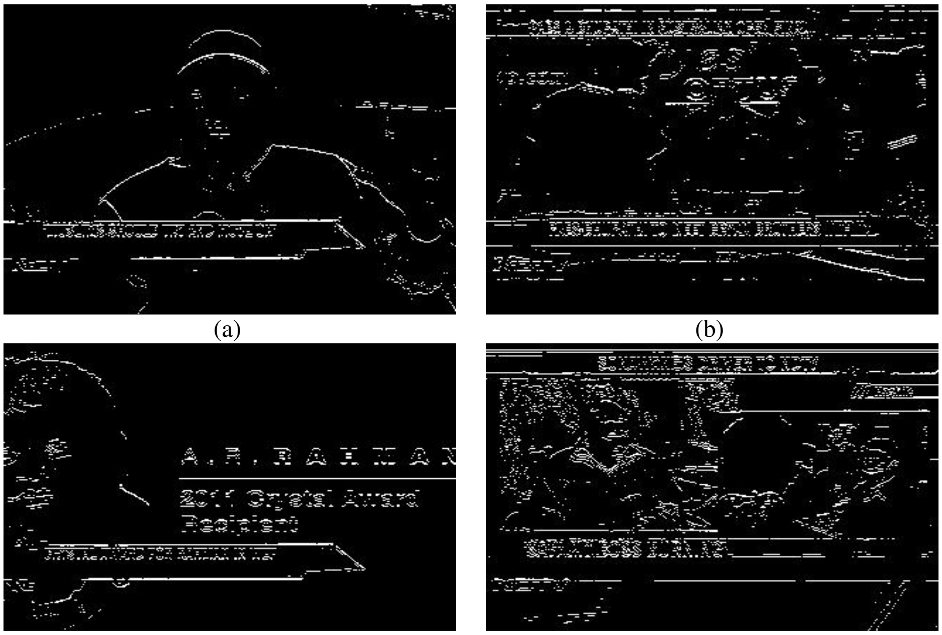

'MUSLIMS SHOULD TRY AND MOVE ON'
NDTV
PAES & BHUPATHI IN AUSTRALIAN OPEN FINAL
yes'
OPTUS
PAES-BHUPATHI TO MEET BRYAN BROTHERS IN FINAL
NDTV

(a) (b)


A.R.RAHMAN
2011 Crystal Award Recipient
CRYSTAL AWARD FOR RAHMAN IN WEF
NDTV
SONAWANE'S DRIVER TO NDTV
Nashik
'SAW MY BOSS BURNING'
NDTV

(c) (d)

Figure 4 Original Frames from News Videos



(a) (b)

(c) (d)

Figure 5 Transition Map Generations for Figure 4

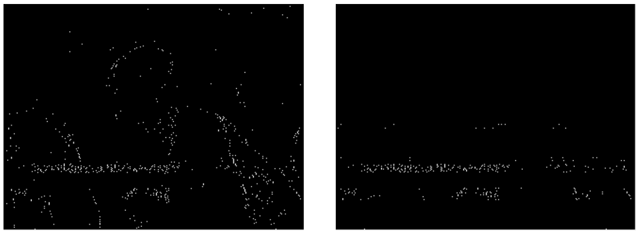
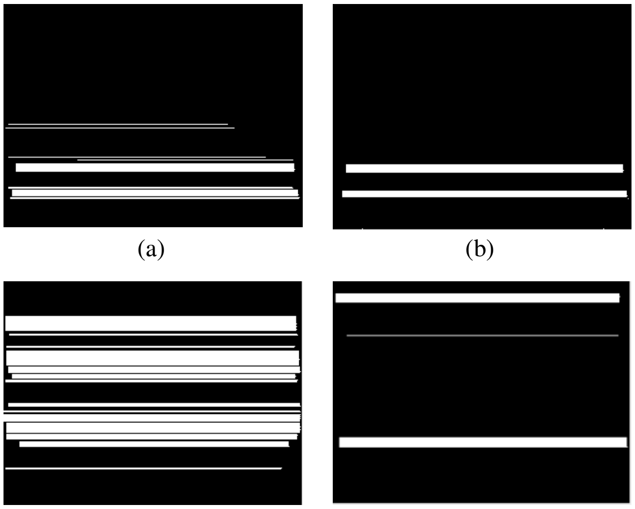


Figure 6 Building Linked Map through Connected Component



(a) (b)

(c) (d)

Figure 7 Overlay text region determination and refinement

5. EXPERIMENTAL RESULTS

NDTV News dataset is used. The project work is implemented in Visual Studio 10 (C# .NET). FFMPEG Library is used to extract the frames form news videos. Overlay scene is also inserted on the video scenelike the overlay text is, the transition region is also observed atthe boundary of the overlay scene. Moreover the boundary ofoverlay scene is vertically long in the transition mapwe have easily removed the boundary of overlayscene from the transition map by the length of connectivity inthe vertical direction. Transition map for Original images in Figure 4 are shown in Figure 5. The results of building linked map are shown in Figure 6. Figure 7 shows the results of boundary smoothed region detected and the refinements made.

CONCLUSION

The approach of transition map to detect the text region is proved to be promising. The work is carried on dataset of NDTV news and large data set of video images containing artificial text developed at Image Processing Center (IPC) a research facility at - National University of Sciences and Technology (NUST), Pakistan. Although this is very specific, the approach is suitable for all generic videos containing text information. It has been observed that such information is available with movies, TV shows uploaded on YOUTUBE, and personal videos. Some of the technologies are getting matured such as speech recognition; face detection, soft computing etc. The “multiple concept detectors” will provide the automatic labels for videos getting uploaded on data network. This will ensure the better retrieval results.

REFERENCES

[1] Agnihotri L. and Dimitrova N., “Text detection for video analysis”, in Proc. IEEE Int.Workshop on Content-Based Access of Image and Video Libraries, Jun. 1999, pp. 109–113.

[2] Bertini M., Colombo C., and Bimbo A. D., “Automatic caption localization in videos using salient points”, in Proc. Int. Conf. Multimedia and Expo, Aug. 2001, pp. 68–71.

[3] Chen Datong, “Text Detection and Recognition in Images and Video Sequences”, PhD Thesis, the Swiss Federal Institute of Technology Lausanne (EPFL), Switzerland, 2003

[4] Chen Jau-Yuen, TaskiranCuneyt, Alberto Albiol, Edward J. Delp, and Charles A. Bouman, “Vibe: A Compressed Video Database Structured for Active Browsing and Search”, Purdue University, 1990

[5] D. Chen and J. Luettin, “A survey of text detection and recognition in images and videos”, RR-00-38, IDIAP, Aug. 2000.

[6] Irfanullah, NidaAslam, Kok-Keong Loo and Roohullah “Semantic Multimedia Annotation: Text Analysis”, in International Journal of Digital Content Technology and its Applications Volume 3, Number 2, June 2009.

[7] J. Cho, S. Jeong, and B. Choi, “News video retrieval using automatic indexing of korean closed-caption,” Lecture Notes in Computer Science, vol. 2945, pp. 694–703, Aug. 2004.

[8] Kim Wonjun and Kim Changick “A New Approach for Overlay Text Detection and Extraction from Complex Video Scene”, IEEE Transaction on Image Processing, Volume 18, Issue 2, Feb. 2009 Page(s): 401 – 411

[9] Otsu N., “A threshold selection method from gray-level histograms”, IEEE Trans. Syst., Man, Cybern., vol. 9, no. 1, pp. 62–66, Mar. 1979.

[10] Pravin M. Kamde, Dr. S. P. Algur, “A Survey on Web Multimedia Mining” Published in The International Journal of Multimedia & Its Applications (IJMA) Vol.3, No.3, August 2011, ISSN 0975-5578.

[11] Sato T., Kanade T., Hughes E. K. Hughes, and Smith M. A., “Video OCR for digital news archive”, in Proc. IEEE International Workshop on Con-tent-Based Access of Image and Video Libraries, Jan. 1998, pp. 52–60.

[12] S. Shiravale , Pravin M. Kamde, “Video OCR for Video Indexing” in IACSIT International Journal of Engineering and Technology (IJET), Vol. 3. No. 3, pp 287-289, June 2011, ISSN 1793-8244. 2011.

[13] Pravin M. Kamde, Dr. P. J. Kulkarni, “Minimum Vertex Cover Problem Optimization Using Hopfield Neural Network” in the International Journal on “Technology, Knowledge & Society, Common Ground Publisher, Australia, published in August 2006 issue.

[14] Jiping Liu; Yanxiang He; Min Peng; , "NewsBR: a content-based news video browsing and retrieval system," Computer and Information Technology, 2004. CIT '04. The Fourth International Conference on , vol., no., pp. 857- 862, 14-16 Sept. 2004.