Scientific Journal of Impact Factor (SJIF): 4.72

e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 4, Issue 5, May -2017

Hybrid approach for news video mining through visual content detection

[1]Trupti N. Visani, [2]Devangi L. Kotak

[1]*Research Scholar,* [2]*Assistant Professor*
[1]*Computer Engineering Department,*
[1,2]*V.V.P. Engineering College, Rajkot, India*

Abstract —*As the expansion in utilization of the Internet, the information exchanged, perused is evolving. We show a framework created for substance based news video perusing for home clients. On the off chance that client has no time for watching news for an entire day then he or she can see highlights which came in TV yet their timings are additionally settle their might be chances client may not ready to see that so for that news synopsis is valuable. There are three fundamental variables that recognize our work from other comparable ones. In the first place, we have incorporated the picture and sound examination brings about distinguishing news fragments and for that we are discovering shots and keys outline extraction for decreasing the time, Second, we utilize the video OCR innovation to recognize content from casings, which gives a decent wellspring of printed data for story order when transcripts and close inscriptions are not accessible. Third thing we are utilizing logo coordinating for break location and breaking news recognition. From proposed strategy client can get customized news rundown*

Keywords*-News extraction, Text detection and reorganization, ad detection, key frame extraction, Edge Detection, text localization, anchor detection, anchors reorganization.*

I. INTRODUCTION

A Both the deluge of the video contents through media such as digital broadcasting and the emerging industry involving PVR, EPG, and large-size storage, are changing the paradigm of how to watch TV. For example, Tivo and ReplayTV are already changing the life style such that viewers can watch any contents in anytime on behalf of viewer's will. The nutshell of this trend is that viewers can record any broadcasting contents through EPG and watch them in a digested or arranged format. Since sports and news programs retain well-organized structures and take much time to watch after recording, there are substantial needs from viewers to see some highlights or storyboards for quick browsing. In the case of the news program, viewers often want to see the main topics and their reports with a storyboard and quickly decide what they want to see more in detail. The anchorperson shot detection has been a fundamental research issue to compose such a news storyboard. There have been active research efforts going on the anchorperson shot detection. The template based search and match method [1], [2], [3] is good for the formatted news structure but has a weakness on news structure or format change. The occurrence frequency and time constraint of the anchorperson shot were used for the anchorperson shot candidate extraction[4]. The motion information was used for the false alarm (e.g. interview shot and report shot) removal.

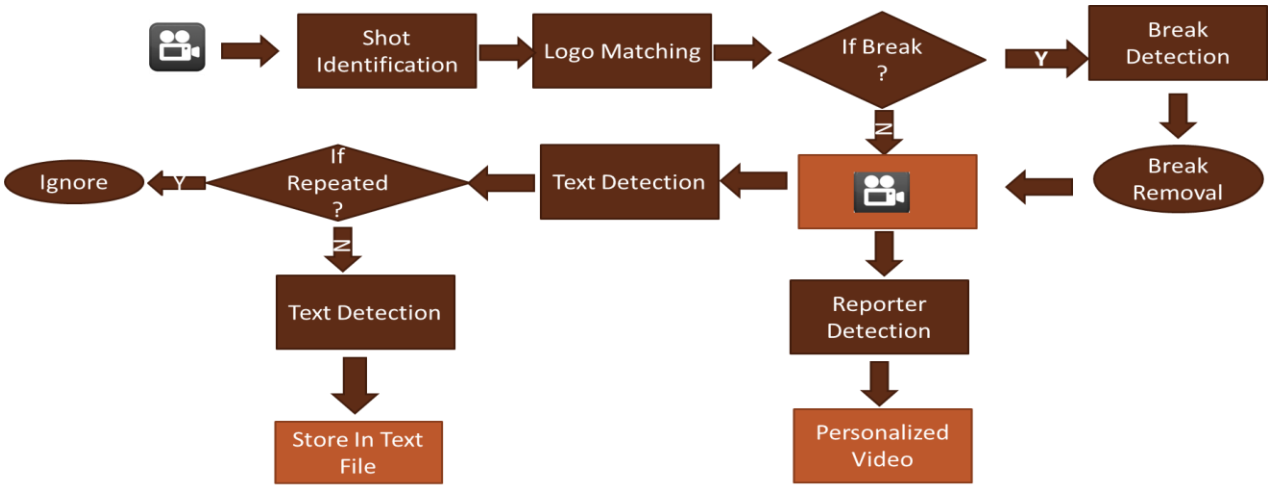
Clustering method [5] was proposed for the similar anchorperson shot clustering in an unsupervised and model-free manner. Multi-feature information such as motion, color, face, cloth, and caption was used in [6]. This extracts faces in DC domain and uses luminance variation in the cloth area to confirm the anchorperson shot. Most of the previous works have been suffered from false alarms caused by anchor person shot like report shots and news format changes.

II. RELATED WORK

In Among all the domain-specific applications, news video processing is probably the most extensively studied topic. Broadcast news is valuable to data analysts in the government, information providers, and television consumers [8]. However, since news events happen daily all over the world, a person cannot afford to view all news shows on all channels indiscriminately. To alleviate the problem, a news video database that compresses and digitally stores news broadcasts and provides interactive retrieval interface over a network needs to be developed. This would enable automatic retrieval of relevant news stories from all the networks and news sources covering the topic of interest [10]. In recent years, several news video processing systems have been developed, such as the MEDUSA system [9], the Broadcast News Navigator System [8], [2], [4], [7] and the Informedia Project [11], [12], [13], [7]–[14]. Most of these systems allow automatic or semi-automatic parsing and annotation of news recordings for interactive news navigation, content-based retrieval and news-on-demand applications. For effective news browsing and retrieval, reliable news story parsing is crucial in the video library system. Correct parsing of news stories leads to much more effective retrieval than simple linear or keyword only search. Merlino et al. [27] empirically demonstrate that the speed with which a user can retrieve relevant stories that are well segmented can be orders of magnitude faster than the speed of linear search or simple keyword-based search. Without good story parsing, all other components of a video library are significantly less useful, because the user cannot locate desired material efficiently

In [14], based on the observation of news programs aired by the Singapore Broadcasting Corporation (SBC), Zhang et al. assume a straightforward temporal syntax of a news video—a sequence of news stories interleaved with commercials. Each news story is composed of an anchorperson shot followed by relevant news footage. Therefore, a news program can be represented as a hierarchical tree-structured model as shown in Fig. 1. Such a simple news structure has been observed for news programs in Singapore (SBC) [2], [4], Hong Kong (ATV, TVB) [12], [6], and local news in the U.S. [15]. News story parsing, therefore, becomes a problem of how to distinguish anchorperson shots from news footage shots. Even for more complicated news program structures, anchorperson shots still serve as the root shots for constructing news stories [12]. Thus, detecting anchorperson shots plays a key role in news story parsing. Most of the existing anchorperson detection methods are based on the model matching strategy [3], [11], [13], [14]. Following Swanberg's proposal [3], Zhang et al. construct three models for an anchorperson shot: shot, frame, and region,
[4]. An anchorperson shot is modeled as a sequence of frame models and a frame is modeled as a spatial arrangement of regions. Thus, recognizing an anchorperson shot involves testing that every frame satisfies a frame model, which in turn means testing each frame against a set of region models. These models vary for different TV stations. It is difficult to construct all the possible models for different news videos. Moreover, the model matching method has a high computational complexity. Gunsel et al. identify anchorperson shots by color classification and template matching [5]. They first extract possible regions where anchorpersons may be situated with skin detection and histogram intersection, then compare these regions with templates stored in the application data. Like Zhang's proposal, the creation and matching of templates are time-consuming processes and strongly depend on the application data. The face detection approach by Avrithis et al. [3] is quite complicated, with a number of parameters needing to be manually tuned. Given that face detection in a still image is already difficult enough, face detection in video is too time-consuming for practical application. Furthermore, since the method also requires training data for classification, it is not an unsupervised approach. The template based method by Hanjalic et al. [8] assumes that different anchorperson models have the same background. This is not true for most news stations. Because of different camera angles, different models have different backgrounds. Adding the changing clothes color and news icon, we find less than 30% similarity between different models in our dataset, which is not enough to distinguish the anchorperson shots. In addition, the complicated template matching method is very time consuming.

III. PROPOSED ALGORITHM



I. Explanation of Proposed Framework

From the literature, I have proposed my system having following 12 steps for effectively summarization of news video.

Step 1: Input video

Take a video as input.

Step 2: Convert video into frames

The input video is converted into the number of frames.

Step 3: Key Frame Generation

Key frames are those frames that best reflect the shot contents to represent the shot. The extracted key frames should covers as much similar content of the shot as possible and keep out as much redundancy as possible. The features used for key frame extraction can include colours (particularly the colour histogram), edges, shapes, optical flow, MPEG motion descriptors, motion vectors etc.

Psudocode: Key Frame Generation

Read all the frames.
For each frame do We have taken Key Frame from every shot 1st , last and middle Frame

Step 4: Shot Boundary Detection

Shot boundary detection for different transitions. First input will be news video then it will transfer to shot Boundary detection then in that cut fade and dissolve detection will be done which helps to identify the shots in movie. In each shot we extract 3 frames for key frame extraction.

In every shot 1st frame last frame and middle frame will be consider as key frames. Those all key frames are database for retrieval phase. So for better retrieval it is needed that there should be good shot boundary detection. Identification of exact cut fade and dissolve will give better retrieval.

So we will combine these all n make one new parameter for detection. Correlation is used for matching process which done sequentially frame by frame.

Color changes will not occur rapidly within a shot. It will only occur between consecutive shots. So only color feature can be used for cut detection. If color histograms between two consecutive frames are greater than threshold then there will be shot change occur.

There is possibility that there will be no same edge between 1st and 2nd frame. The number of input and output edges can be counted[3]. If both the frames belong to the same shot then the number of input and output edges is more or less same. So if difference is very large then shot change can be occurs. If there are n pixels in the frame and the in edges are Xin and the out edges are Xout for frames f and $f-1$ respectively, then the edge change ratio (ECR).

Fade and Dissolve transitions are combination of repetitive change in pixel intensities. Using sliding window one can observe frames for their intensity change. If adequate numbers of pixels show repetitive change in their intensity within the window, we can assign current frame to be frame in dissolve transition [12].

We are using single feature that can be use for every transition. That feature combines different methods for cut fade and other gradual transition.

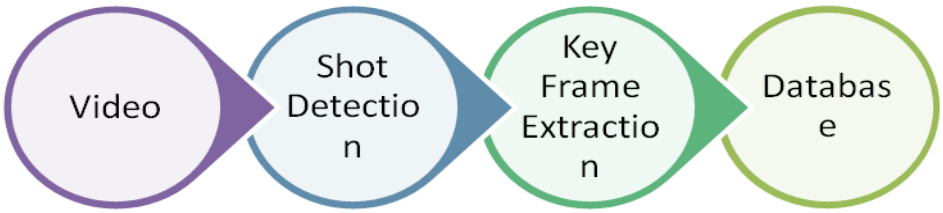


Fig 3.2.1 Proposed Framework for Shot boundary detection

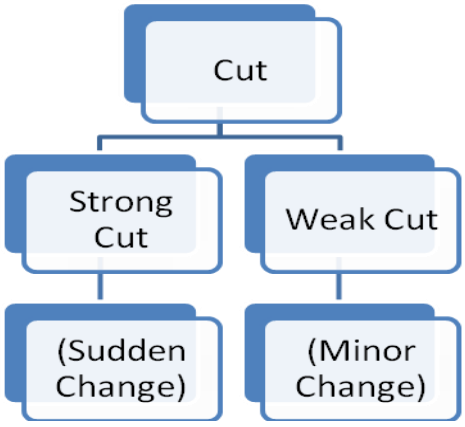


Fig 3.2.2 Cut Detection

Algorithm for shot detection

Step-1: Frame Conversion
Step-2: Computer HSV histogram
Step-3: Find Discrete Cosine transformation
Step-4: Compute edge detection
Step-5: Find entropy
Step-6: Calculate standard deviation
Step-7: Go to step-2 while end of the frames
Step-8: Correlation of parameters
Steo-9: Find unique parameter with combination of this all features. New parameter= 1 / (hsvhistocorr+graycorr+histocorr+edgecorr)

Step-10: Find Mean and Standard deviation of new parameter
Step-11: Calculate Thresholding Tb = Mean + a * standard deviation where a is constant
Step-12: Start from 1st frame
Step-13: If new parameter of frame > Tb Strong cut detected Else If new parameter of frame > (mean + std) Weak cut detected
Step-13 If entropy of frame is zero Fade detected
Step-14 If accumulated difference of frame > Tb Gradual transition detected
Step-15 Go to step-13 while end of the frames

Step 5: Store reference logo frame of channel

The next step is to give the logo frames same as the logos used in particular news.

Step 6: Compare logo frame for break detection

In this step, the logo frame is compared with the key frames. The number of starting logo frame and ending logo frame is stored.

In each and every break there is an logo frame at starting and ending..

If Logo Then its break

Algorithm for logo matching

Read all logo frames.
For each logo frames do
Read all key frames.
For each logo frames do
Convert RGB image to GRAY scale
Compute grey image histogram
Compute correlation coefficient of logo frame and key frame
If (correlation coefficient > threshold) then Store image

Step 7: Removal of break Frames

In this step, frames are extracted from the result of previous step. These frames are the break frames.

Remaining frames will be of news frames.

Step 9: Output video of news

In this step from the database of extracted frames video will be generated.

Step 10: Anchor detection

In this step from key frames If face in any frame? Then crop Match with database of anchor
If match? Then labeled
Else Go to step 12

Step 11: Text Extraction from frames

In this step text first detected and then recognized from image. If it is not clear then it will be enhanced. If repeated text then ignored and then final breaking news will be stored in document.

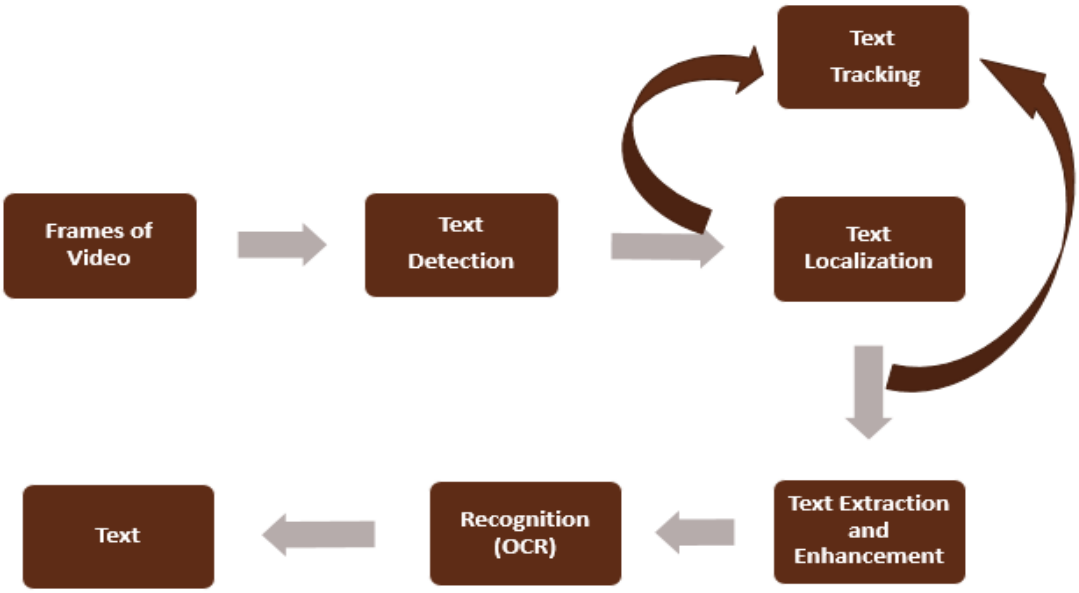


Fig 3.2.3 Framework for text extraction

Step 12: Output Text

In this step all text will be store in document and will give output of document which contains breaking news.

II. Result Analysis

Shot Detection:

Figure 4.1.1 Cut detection.

Figure 4.1.2 Fade detection.

Figure 4.1.3 Dissolve detection.

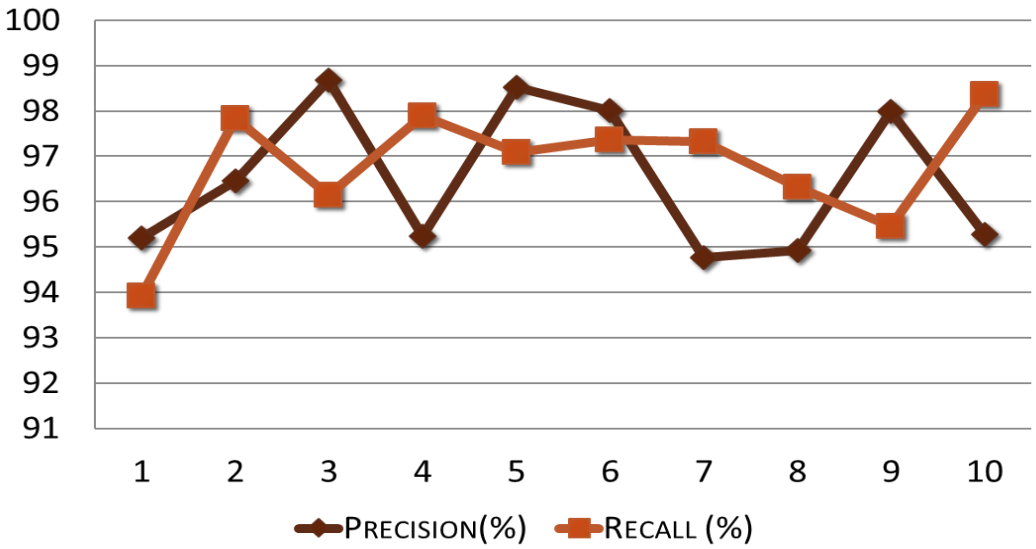


Figure 4.1.4 Analysis of shot detection

Key frame extraction:

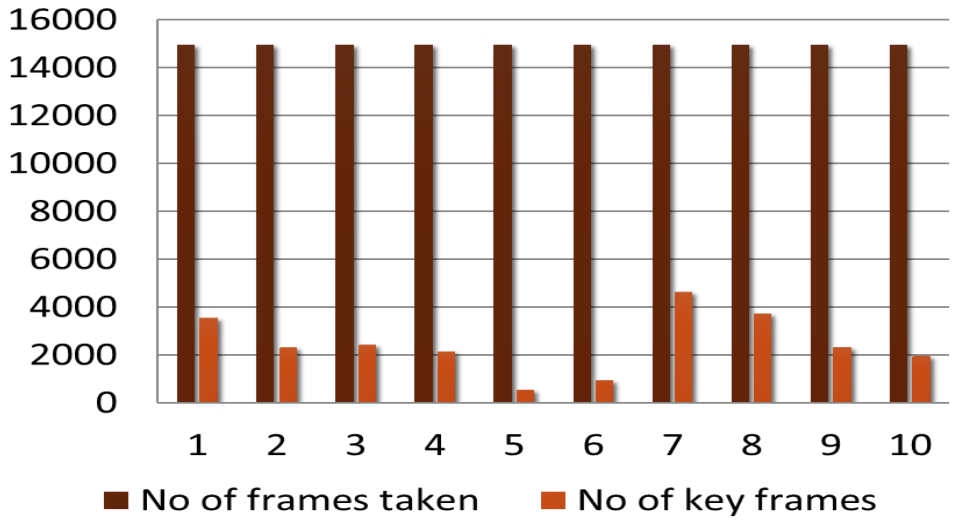


Figure 4.1.2.1 Analysis of Key frame extraction

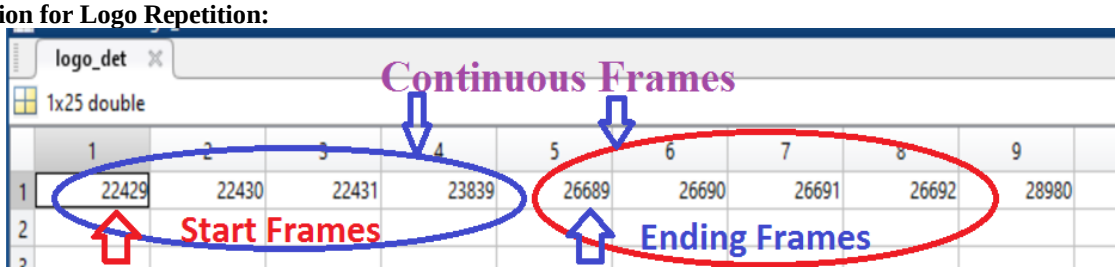
Logo Matching:

Figure 4.1.3.1 Some Reference Logo Frames

For Logo Repetition:

Figure 4.1.3.2 Repetition of Logo Frames

Solution for Logo Repetition:



Step-1 : Keep 1st Logo Frame Discard Remaining But There Can be Only Starting Logo Frame No Ending Frame Then
Step-2 :We have taken Starting Logo -> Next Strong Cut
Step-3 : From Final Starting and ending replay we have selected Frames between -> 1st & 2nd -> 3rd & 4th detected logo's

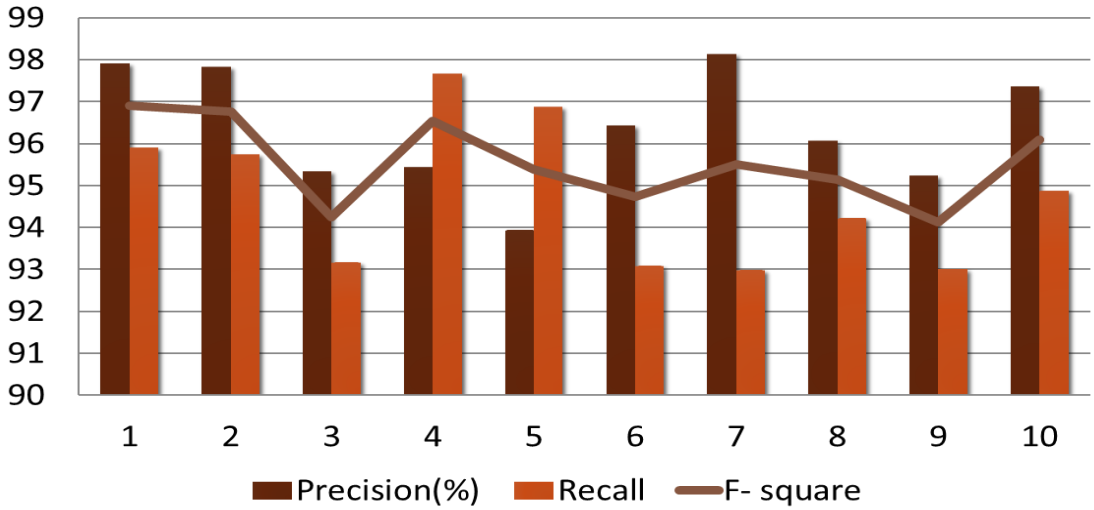


Figure 4.1.4.1 Result Analysis for logo matching

Face Detection:

Figure 4.1.4.2 Some Reference Logo Frames

III. CONCLUSION

From the literature, we observed the various methods used in news content extraction. We addressed the issue of summarization and description of news videos. It's shown that to reduce the time consumption scene recognition and key frame extraction is been used and from key frames break's logo is matching and which is very effectively done. We are getting very good results for break extraction from news video. After removal of break if user need anchor based news then for that we can do highlight extraction based on reporter and if user need only text then text extraction can be done. So we have implemented text extraction frame by frame. There is lots of information in each frame so lots of work can be done in this area.

I. REFERENCES

[1] Kumar, Punith, and P. S. Puttaswamy. "Moving text line detection and extraction in TV video frames." Advance Computing Conference (IACC), 2015 IEEE International. IEEE, 2015.
[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
[3] Covell, Michele, Shumeet Baluja, and Michael Fink. "Advertisement detection and replacement using acoustic and visual repetition." 2006 IEEE Workshop on Multimedia Signal Processing. IEEE, 2006.
[4] Jung, Cheolkon, and Licheng Jiao. "Korean-English bilingual videotext recognition for news headline generation based on a split-merge strategy." Journal of Real-Time Image Processing 11.1 (2016): 167-177.
[5] Wang, Jinqiao, et al. "Automatic TV logo detection, tracking and removal in broadcast video." International Conference on Multimedia Modeling. Springer Berlin Heidelberg, 2007.
[6] Kannao, Raghvendra, and Prithwijit Guha. "TV News Commercials Detection using Success based Locally Weighted Kernel Combination." arXiv preprint arXiv:1507.01209 (2015).
[7] 10Qi, Wei, et al. Integrating visual, audio and text analysis for news video. Image Processing, 2000. Proceedings. 2000 International Conference on. Vol. 3. IEEE, 2000.
[8] 4Kumar, Anubhav. An Efficient Approach for Text Extraction in Images and Video Frames Using Gabor Filter. International Journal of Computer and Electrical Engineering 6.4 (2014) 316.
[9] Montgomery, Martin, and Debing Feng. "'Coming up next': The discourse of television news headlines." Discourse & Communication (2016): 1750481316659397.
[10] Thai V. Hoang , S. Tabbone(2010),"Text Extraction From Graphical Document Images Using Sparse Representation"in Proc. Das, pp 143–150.
[11] Sumathi, C. P., T. Santhanam, and G. Gayathri Devi. "A survey on various approaches of text extraction in images." International Journal of Computer Science and Engineering Survey 3.4 (2012): 27.
[12] Davod Zaravi, Habib Rostami, Alireza Malahzaheh, S.S Mortazavi(2011)," Journals Subheadlines Text Extraction Using Wavelet Thresholding And New Projection Profile", WorldAcademy Of Science, Engineering And Technology .Issue 73.
[13] Syed Saqib Bukhari , Thomas M. Breuel,Faisal Shafait(2009), "Textline Information Extraction From Grayscale Camera-Captured Document Images ", ICIP Proceedings Of The 16th IEEE International Conference On Image Processing, pp: 2013 – 2016.
[14] Kohei Arai1 , Herman Tolle(2011)," Text Extraction From Tv Commercial Using Blob Extraction Method", International Journal Of Research And Reviews In Computer Science Vol. 2, No. 3
[15] Wonder Alexandre Luz Alves And Ronaldo Fumio Hashimoto(2010),"Text Regions Extracted From Scene Images By Ultimate Attribute Opening And Decision Tree Classification", Proceedings of the 23rd Sibgrapi Conference On Graphics, Patterns And Images.
[16] Gllavata, Julinda, Ralph Ewerth, and Bernd Freisleben. "A robust algorithm for text detection in images." Image and Signal Processing and Analysis, 2003. ISPA 2003. Proceedings of the 3rd International Symposium on. Vol. 2. IEEE, 2003.
[17] V.Vijayakumar,R.Nedunchezhianm(2011),"A Novel Method For Super Imposed Text Extraction In A Sports Video",International Journal Of Computer Applications,Volume 15– No.1.
[18] Trung Quy Phan ,Palaiahnakote ,Shivakumara Chew Lim Tan(2010)," A Skeleton-BasedMethod For Multi-Oriented Video Text Detection", Das '10 Proceedings Of The 9th IAPR International Workshop On Document Analysis Systems,pp :271-278.
[19] Gao, Xinbo, and Xiaoou Tang. "Unsupervised video-shot segmentation and model-free anchorperson detection for news video story parsing." IEEE Transactions on Circuits and Systems for Video Technology 12.9 (2002): 765-776

