Unsupervised Ads Detection in TV Transmissions

Waseemullah[1], Najeed Ahmed Khan[2]
Computer Science & IT Department
NED University of Engineering & Technology
Karachi, Pakistan

Umair Amin[3]
We4do Solutions Karachi
Pakistan

***Abstract*—A novel framework is presented that can segment semantic videos and detect commercials (ads) in a broadcasted TV transmission. The proposed technique combines SURF features and color Histogram in a weighted combination framework resulting in detecting individual TV ads from the transmission after segmenting semantic videos. Thus, better results are achieved. The proposed framework is designed for TV transmissions those who do not use black frame technique between the ad and non-ad part of the transmission and is commonly used in Pakistani TV channels transmission. The television transmission standards in Pakistan are different from those that are used in other countries of the world. The framework used unsupervised technique to segment the semantic videos.**

Keywords*—TV ads; video segmentation; semantic analysis; ad segmentation; unsupervised segmentation

I. INTRODUCTION

Semantic analyses involve the extraction or understand the meaning of the subject that is being analyzed and finding relations among unstructured data. It is most commonly used in analyzing text data and Natural Language processing system.

The amount of video information that is being generated now is more than it has been ever before. From TV transmission to CCTV footage and videos made by cell phones everything is adding to the amount of video data present. Therefore, it is important to have tools that can automatically extract meaningful information from these videos.

Advertisements displayed in broadcast TV transmissions are a very important part of transmission as major revenue for a broadcaster is generated by advertising. However, the advertiser is keen in statistics of the broadcasted ad in the transmission. Knowing who what and where advertising can be useful information to measure the market trends and forming business strategy.

If this statistics computed automatically and reliably then things will easy for both the broadcaster and advertiser as well for the market trends analyzer.

The term semantics is a broad term and can be used to measure similarity in different domains such as sports, scenery, image etc. The proposed framework chose for the automatic ads detection that appears in a broadcasted television transmission. The developed framework can differentiate between ads and non-ad part in video of transmission as well as compute the statistics of broadcasted an ad in the transmission. Several research works have been attempted commercials detection and identification however; follow the USA and European television transmission standard which follow black frame transmission technique [1]. There are TV channels transmission e.g. in Pakistan that do not follow appearance of a black frame between the ad and a non-ad part of the transmission.

The developed framework useable for TV transmissions those do not use black frames technique in between ad and non-ad video part in the transmission like in Pakistan TV transmission. Techniques are common such as in [1] use of a black frame between ad and non-ad part in video to detect boundary of an ad. Another technique to detect boundary of an ad in video part is the nonappearance of TV channel logo. During the commercial breaks the TV channel logo is removed that to commercials video part and non-commercial video part in TV transmission [2]. This technique also does not use in Pakistani TV transmissions. Fig. 1 shows the some examples labeled in the dataset of ads that contains different products images.

This paper contains 6 sections including this section. Section I described Introduction of framework. Section II described related work that has been done in relation to video segmentation and semantic analysis for ad detection. Section III described implementation of framework. Section IV presented results of experiments that were conducted. Section V presented conclusion and future work.

II. LITERATURE REVIEW

Most American and European television broadcasters often use the black frame as standard for the transmission for separation of TV advertisements from the other programs. Many researchers have worked on TV adds detection using the black frame in TV transmission (Fig. 1).

Fig. 1. Examples (a-l) from the dataset of ads contain different products images/information.

In [2] L. Y. Duan et al. worked on hassle of recognizing and categorizing TV ads from TV video streams. For boundary detection in each commercial a multimodal approach is followed. Absence of TV channel logo is used for separating a normal transmission program from a TV ad and also used for segmentation of ads. In addition to that Hidden Markov Model was trained using audio and visual feature for TV ad segmentation. Results showed the precision and recall were 90% and 80% respectively.

In research findings [3] Covell M. et al. investigated problem of extracting video fragments that are appeared frequently in video streams. Their real goal was the identification of those TV ads those are identified with telecasters possess programs, for example, programs that will be played one week from now in a single day at an alternate time. Their suggested approach has three noteworthy advances sound redundancy identification, endpoint discovery and discovery of endpoint. Experiment ran video recording of four days on video taken of different TV channels. An precision and recall was achieved of 85% and 94% respectively for sound coordinating part. Precision and recall was 92% and 93% achieved after matching video streams announced and finally outcome values for precision and recall were 99% and 95% respectively.

In [4] the issue of recognizing TV ads in MPEG compressed video footage is investigated. It uses features extracted using MPEG parameters. TV ads detection is achieved using black frame, unicolor frame and in addition to that changed in aspect ratio. It was additionally watched that one minute was length for each TV ads break. For evaluation eight hour length video stream was recorded and appearance of black frame in TV ads break was observed to be the most appropriate advertisement recognizing parameter.

In [5] issue of accessing TV Ads by considering salient semantics is addressed through semiotic point of view. Critical, playful, utopic and practical were recognized as semiotic classes. 150 ads recorded from Italian TV networks and examined to check that groupings made by framework are same as obtained via human specialists. For playful ad highest accurate results were obtained for practical least accurate results were obtained due to wrongly recognizing product in foreground.

In [6] the video contents were analyzed for real time indexing. The proposed achieves video indexing by taken into account the video contents using Hidden Markov Model (HMM). The HMM accepts input feature that is based on image difference that indicates location target object in shot. In addition to that grey level histogram, intensity of motion and average motion deviation features are used in this approach. For assessment videos from 12 news programs obtained from various German TV networks were examined. Half of them were used in training and half used in testing. Nine different classes were recognized. Seven from total of nine had detection rate more noteworthy than 80%. The experimental showed that detection value for short news is higher than long news programs.

El-Khoury et al. [7] suggested a strategy for automatic extraction from TV content based on signals. His technique can be applied on sound, visual contents or a mix of sound and visual contents by making utilization of Bayesian Information Criterion and general likelihood ration. This framework was assessed on recorded videos from French TV and furthermore using TRECVid dataset. Recall value of 93% and 89% was observed for ARGO and TRCVid respectively whereas precision value of 93% and 91% was examined for ARGO and TRECVid respectively.

In [8] X. S. Hua et al. suggested a learning method for recognition of TV advertisements. His proposed method performs classification through Support Vector Machine on the basis of audio and video features. For visual features variance and average of frame difference and variance of edge change ratio were used. For audio Mel frequency Cepstral Coefficient and shot time energy were taken into account. For the purpose of evacuating of shots that have little duration, checking of long ads and refining TV ads limits. For assessment TV transmission recordings of duration10.75 hours used gathered from various TV channels including: NBC, CNN and ESPN2. Recall value 88.21% and precision value 89.39% seen without post preparing and after post handling, it expanded to 91.65%.

In [9] S. C. Chen discussed the issue of identification of scene transformation in video streams by utilization of sound and pictorial data that is accessible in video. The suggested strategy comprises of first deciding shot limits and then unsupervised segmentation using object tracking. During experiment many video recordings were made from different TV stations. The results reported after experiments were 89% and 92% for the average recall and average precision respectively.

In [11] the issue of identifying feature films in different types in light of their preview is investigated. They have grouped movies into four types called Action, Drama, Horror and Comedy depending on visual cues. The proposed

approach comprised on group of input features utilized to lessen variance among points of same class and maximize difference among points of other class. They computed average shot length feature in Hue, Saturation, and Value of brightness space. For experiment more than hundred previews of movies were downloaded from Mac website and grouped in four classes. There experiments on 101 videos showed 17 video incorrectly categorized.

In [12] K. Schöffmann et al. focused on the issue of detecting TV ads in recordings coded in H.264/AVC. This approach is different from other existing approaches as it works straight on compressed streams as opposed to uncompressing video. The proposed approach makes utilization unavailability of TV station at the time of TV ads which is valid for European and especially German TV. For assessment of results recording from 19 TV stations was utilized that were prevalently related German watchers. Results showed that recall value and precision value were reported as 97.33% and 99.31% respectively.

In [13] J.R. Smith et al. the issue of automatically marking on TV transmission for later pursuit and indexing is addressed because annotating by human is costly and tedious job. For audio, text and video analysis a multimodal approach is applied. Their framework combines the text, audio and video data for video stream annotation. Using machine learning techniques semantic models library is developed using training dataset. Little human intervention involved for small portion of dataset in training stage. Framework enabled clients to search videos either by feature, choosing key frames, semantically and via text. With the end goal of assessment TRECVID benchmark is used. 90% precision was obtained.

N. Venkatesh et al. [14] suggested a novel technique for TV commercial detection based on cookery programs. TV ads are identified using audiovisual features appeared in transmission. At first sound features are utilized for recognizing begin and end of TV ad break. At that point program name logo is compared with begin and end of commercial break. The audio features i.e. short time energy as well as Zero crossing rate are considered. Visual analysis is carried through edge identification and corner.

In [15], Bingqing Qu et al. investigated the content based segmentation of TV programs through grammatical inference for the sake of video indexing. They proposed symbolic approach that finds a common structure model for repeating programs. Each structural element is having distinct syntactic meaning from other video segments. This distinct structure is represented by a symbol and used in grammatical inference. Their defined approach needs minimal information for segmentation of TV video streams which is based on two steps. In first step symbols are assigned to similar structures of programs. In second step, for searching particular TV program grammatical inference is taken into account using symbolic representation of each structural element. The experiment was performed on four distinct types of repeating programs like talk show, news, magazine and game. The results were obtained using three methods called base line, multiple sequence alignment (MSA) and uniform resampling (UNR) methods. UNR was better than baseline in results whereas MSA out performed UNR in situation where program duration is much lengthy. Precision and recall value for news, talk show and games and magazine were 69% and 54%, 53% and 54%, 69% and 50%, 42% and 64% respectively for MSA method.

In [16] Xiaomeng Wu, et al. suggested a system based on detecting and locating a TV commercial in video. They used Hashing algorithm for detection and localization of commercial. The system performance claimed consistent for both video and audio, high speed and clustering is completely unsupervised. Extraction is improved by algorithm that is responsible for integrating audio and video footage. Accuracies were 97.4% and 98.1% for frame level and sequence level during testing. This system is based on new BOFs model as it smart enough to handle decoding errors. The system is tested on ten hours, a month and five year video effectively and efficiently.

III. FRAMEWORK

The proposed framework starts with the add detection in TV recorded transmission that contains recording of different programs including ads (commercials). This section address the approach used for detection of ads in video if present.

A. Ad Break Detection

To segment the videos automatically into ad and non-ad, we devised an algorithm. It calculates variance of a window before and after a point. The difference in variance between left window and right window is greater at the beginning and end of ad-break. So it can be used to segment video between ad and non-ad parts.

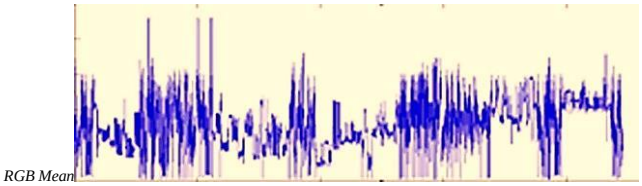

Fig. 2. Plot of RGB for each frame Vs RGB mean where ad breaks had higher distortion in graph and the portion that corresponds to normal transmission had fewer peaks per second.

In the first step for the frames in video RGB mean is calculated and plot them along with the corresponding frame numbers. Fig. 2, 3, and 4 show the RGB mean and the frame numbers. Based on the variance shown in the graph, it is possible to see that there was a pattern being followed by ad breaks and normal transmission where ad breaks had higher distortion in graph and the portion that corresponds to normal transmission had fewer peaks per second.

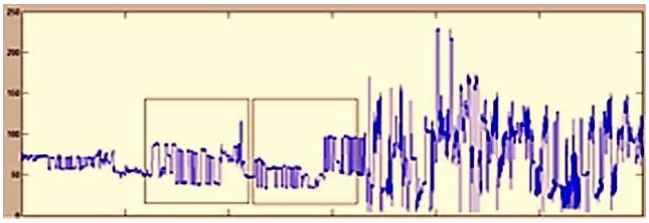

Fig. 3. Left and right window in non_ad region where difference in variance between left window and right window is almost same.

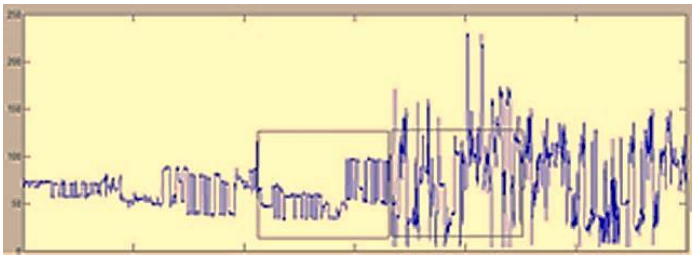

Fig. 4. Left and right window at split region where difference in variance between left window and right window is greater at the beginning and end of ad-break.

In the second step a window size is defined that calculate variance of RGB mean. For each point the variances of that window before that point and after that point is calculated and stores the difference of that variance.

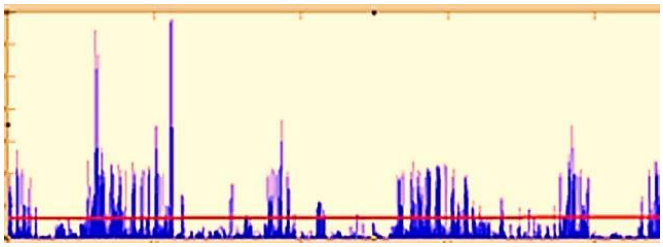

Fig. 5. Plot after step 2 majority of values above red line correspond to ad region using window size 100 frames.

In the next step video is split into ads and non-ad region based on calculating variance, method described above. Main consideration here is that when a value found greater than a threshold value, some other frames seek ahead to make sure that the peak observed is not a false peak. There are several values greater than threshold. Fig. 5 shows add region above the red line after setting a threshold value. Using parameters defined in Table I, we classify values as a sequence of low and high values.

TABLE I. PARAMETERS AND THEIR DESCRIPTION

Parameter	***Description***
Thresh	Threshold to classify a value as low or high
fLowSeek	Number of frames to look ahead when looking from low to high
nLowSeek	Minimum number of high frames that should be present in search space for classification from low to high
fHighSeek	Number of frames to look ahead when going from high to low
nHighSeek	Maximum number of high frames that can be present in search space for it to be classified as change from high to low
fHighSeek2	Minimum number of consecutive low frames for classification from high to low

Parameters and their description were used for classifying values as a sequence of low and high values.

Based on the value of parameters following is established:

- If Value is greater than the threshold mark it as high value if it is lower than threshold mark it as low value.
- If current value is low and previous value was also low mark it as part of same sequence.
- If current value is high and previous was also high mark it as current high sequence.
- If current value is high and previous value was low seek for next “fLowSeek” frames and count number of frames that have value higher than threshold. If that number is greater than “nlowSeek” we mark as a change point from low to high.
- If current value is low and previous value was high we check next “fHighSeek2” frames. If anyone frame in this range is found to be high we consider this value as part of existing high sequence. Otherwise we seek next fHighSeek2 frames, if number of high frames in these is less than nHighSeek we consider this as point of change from high to low sequence otherwise we make this frame as part of existing high sequence and continue.

The proposed Ad Break Detection algorithm was evaluated using the following parameters on a 2.5 hours video recorded from a local TV channel (ARY Digital-Pakistan) at 25 fps.

TABLE II. PARAMETERS AND THEIR CHOSEN WEIGHTS

Parameter	***Value***
Window Size	100
Thresh	800
fLowSeek	2500
nLowSeek	300
fHighSeek	3500
nHighSeek	55
fHighSeek2	800

Parameters and their chosen weights were used for evaluating Ad Break Detection algorithm.

Using these values shown in Table I, the video was split into 15 segments, 7 of them were ad segments and 8 were marked as non-ad. Total length of ad segment was 1 hour 2 minutes. Out of them 4 minute 30 seconds were non-ads that were classified as ads.

Remaining duration was classified as non-ads in them total 4 minutes of content were composed of ads that were wrongly classified as ads.

The (Ad Break Detection algorithm) was again evaluated on altered the values of the parameters in Table II. The altered values are shown in Table III.

TABLE III. PARAMETERS AND THEIR ALTERED WEIGHTS

Parameter	***Value***
Window Size	100
Thresh	700
fLowSeek	2500
nLowSeek	300
fHighSeek	3500
nHighSeek	55
fHighSeek2	2000

Parameters and their altered weights were used for evaluating Ad Break Detection algorithm

Using these values the video was again split into 15 segments, 7 of them were ad segments and 8 were marked as non-ad. Total length of ad segment was 1 hour 4 minutes. Out of them 5 minute 30 seconds were non-ads that were classified as ads.

Remaining duration was classified as non-ads in them total 2 minutes of content were composed of ads that were wrongly classified as ads. This shows choice of parameters values shown in Table II is good as it gives less wrong classification of ads.

B. Ad Boundary Detection

After video segmentation into ad and non-ad parts, it is aim to find automatically the boundaries of individual ads. This task is easy when a black frame exist at the beginning and at the end of each ad part in the video segment. However new techniques are required if black frame not exist a TV transmission, like in Pakistan. Therefore, a novel technique for the ad and non-ad detection automatically in such kind of TV transmission is proposed.

Since the audio was not a good feature to separate ads, as there was no real silence present in between ads. Also several times inside a same ad, block of silence are found that cause wrong attribution of ad boundary. Therefore, instead of audio a new approach based on repeating patterns of scenes was used that assigned ID to each scene and determine ad boundary. The frame work of the proposed technique is shown in Fig. 7.

To detect an ad boundary we employ a method that combines histogram with detecting repeating patterns of scenes in video, as the commercial ads are usually repeated several times during transmission. In the first part of ad boundary detection histogram is calculated for all frame having 30 bins (set threshold value after initial experiments) for each color and scene detection are performed based on histogram. To verify or determine a frame belong to a same scene or not, histogram difference of change point for last 05 scenes is calculated, if appropriate match was found then the frame considered belongs to an existing scene. Fig. 6 shows an output example of different frames belongs to a same scene.

Fig. 6. Frames that marked as belonging to a same scene based on color histogram.

The working of algorithm is divided into two steps.

In first steps scenes are generated from whole video by making use of color histogram. After a scene is identified, it is compared with other scenes that have been already generated from the video. The aim is to assign same scene ID for those scenes which are similar but having not exactly same scene ID. If a new scene appears that was not presented earlier, a new scene ID is assigned. Whereas if that scene already observed at earlier location in video it is added to the scenes list belonging to existing scene ID.

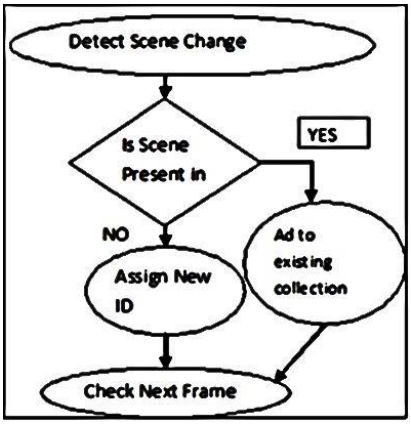

Fig. 7. Proposed framework for boundary detection of an ad.

In second step repeating sequences of scenes are detected. This step allows detecting of unique ads because normally an ad is composed of several scenes and an ad appear several time in a video stream of TV transmission as compared to non-ad transmission. Therefore, if an ad is composed of for example scene number “4, 5, 6” then we find this pattern multiple times in video stream. Thus, if multiple times in video we find that scene 5 comes after scene 4 and scene 6 come from scene 5, it is concluded that these three scenes are related. And because this is a repeating sequence, therefore it is a strong candidate for being an ad.

The details for second step are as follows:

Scenes those have length less than 25 frames are eliminated. Starting from scene that has highest number of occurrences in videos, a repetition threshold is defined as follows:

$$\Theta = (\lambda / 10) \times 3$$

Where, Θ - represents Count Thresh.

λ- Represents Occurrences of pattern.

If Count Thresh is less than three (03) it is increased to three (03). Longest such pattern where number of occurrences of pattern is greater than Count Thresh is searched.

If search is successful then all occurrences of that pattern are marked as checked and added to collection of discovered ads. This process is repeated for remaining scenes by taking unchecked scenes in order of their frequency.

Each discovered pattern generally represents an ad that was repeated several time during the video. First scene of pattern is generally first frame of ad and last scene of ad.

C. Ads segmentation

For Ad segmentation RGB histogram is calculated for all the frames in the video segment and then split ad into scenes by using histogram difference. If histogram difference of two

consecutive frames is greater than threshold the point is considered as a change point for scene.

Inside each scene we define key frames. Key frames are those frames whose histogram difference is greater than the threshold value compared to the previous key frame. Using key frame, each ad is divided into scene with each scene having histograms of all frames and key frames marked inside them. To improve the detection and to eliminate false positives, user asks to select a key object that is used for computation of SURF (Speed Up Robust Transform) feature. The key object should be the main object in the ad, which is significant with respect to the product ad. Example of a selected key object is shown in Fig. 8.

Fig. 8. Key object in an ad for SURF feature computation.

Multiple objects in a same ad can be selected for computation of SURF feature but one object is enough to eliminate false positive ad detections. After this the computed training data can be saved inside a training file that will be used in detection stage.

IV. Results

This section will describe results and outcome statistics of the proposed algorithm evaluated on test dataset. Test dataset composed of 2.5 hours video transmission segment that was recorded from ARY Digital-Pakistan TV channel.

For evaluation of the proposed algorithm, first the ads were marked manually to collect ground truth and then ad detection algorithm was executed on the test dataset segment. To evaluate the performance of the proposed algorithm, automatic labeled outcome results are compared with the ground truth. The target was to find those ads that appear at least 3 times in the provided test video segment. Standard quantitative measures Precision, Recall and F1 value [10] are used to measure the performance of proposed algorithm. Table IV shows the quantitative output values of Precision, Recall and F1 values.

TABLE IV. Precision, Recall and F1 Values

Test Video	*True +ve*	*False + ve*	*False -ve*	*Precision*	*Recall*	*F1-Value*
2.5 hours video transmission ARY Digital-Pakistan	1325	50	116	96%	92%	93%

In Table IV, Columns, True +ve shows number of correct detections duration in seconds, False +ve shows number of incorrect detections duration in seconds, False –ve shows number of missed detections duration in seconds for ads.

Table IV shows Precision; Recall and F-1 Value results evaluated on a Pakistani TV channel (**ARY Digital-Pakistan**) recorded video of duration 2.5 hours. Average precision 96% and average recall rate is 92% and F1 value is 93%.

V. Conclusion and Future Work

The proposed methods incorporate video segmentation techniques based on color Histogram and detect commercials by computing SURF descriptors feature from TV transmission. The experimental results show that a SURF feature provides better segmentation results for the commercials in video.

The framework can be extended for multiple ads detection and identification from which any particular TV commercial can be identified. In addition to commercial identification, statistical analysis can be performed to know about the repeated occurrence of a TV commercial in given video stream. It can also be converted into a complete media monitoring and commercials verification package. This will be suitable for the analysis of commercials in a TV transmission where other solutions are not available.

References

[1] Chen, Shu-Ching, et al. "Scene change detection by audio and video clues." Multimedia and Expo, 2002. ICME'02. Proceedings. 2002 IEEE International Conference on. Vol. 2. IEEE, 2002.

[2] Colombo, Carlo, Alberto Del Bimbo, and Pietro Pala. "Retrieval of commercials by semantic content: the semiotic perspective." Multimedia Tools and Applications 13.1 (2001): 93-118.

[3] Covell, Michele, Shumeet Baluja, and Michael Fink. "Advertisement detection and replacement using acoustic and visual repetition." Multimedia Signal Processing, 2006 IEEE 8th workshop on. IEEE, 2006.

[4] Dimitrova, N., et al. "Real time commercial detection using MPEG features." Proceedings of the 9th International Conference on Information Processing and Management of Uncertainty in Knowlwdge-based Systems (IPMU2002). 2002.

[5] Duan, Ling-Yu, et al. "Segmentation, categorization, and identification of commercial clips from TV streams using multimodal analysis." Proceedings of the 14th ACM international conference on Multimedia. ACM, 2006.

[6] Eickeler, Stefan, Andreas Kosmala, and Gerhard Rigoll. "A new approach to content-based video indexing using hidden markov models." Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS). 1997..

[7] El-Khoury, Elie, Christine Sénac, and Philippe Joly. "Unsupervised segmentation methods of TV contents." International Journal of Digital Multimedia Broadcasting 2010 (2010).

[8] Hua, Xian-Sheng, Lie Lu, and Hong-Jiang Zhang. "Robust learning-based TV commercial detection." Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on. IEEE, 2005.

[9] Lienhart, Rainer, Christoph Kuhmunch, and Wolfgang Effelsberg. "On the detection and recognition of television commercials." Multimedia Computing and Systems' 97. Proceedings., IEEE International Conference on. IEEE, 1997.

[10] Khan, Najeed Ahmed, and David C. Hogg. "Unsupervised Learning of Object Detectors for Everyday Scenes." Training 10 (2014): 11.

[11] Rasheed, Zeeshan, Yaser Sheikh, and Mubarak Shah. "On the use of computable features for film classification." IEEE Transactions on Circuits and Systems for Video Technology15.1 (2005): 52-64.

[12] Schöffmann, Klaus, Mathias Lux, and Laszlo Böszörmenyi. "A novel approach for fast and accurate commercial detection in H. 264/AVC bit streams based on logo identification." International Conference on Multimedia Modeling. Springer, Berlin, Heidelberg, 2009.

[13] Smith, John R., et al. "Learning and classification of semantic concepts in broadcast video." Proceedings of the International Conference of Intelligence Analysis. 2005.

[14] Venkatesh, N., B. Rajeev, and M. Girish Chandra. "Novel TV commercial detection in cookery program videos." Proceedings of the World Congress on Engineering and Computer Science 2009 Vol II, WCECS 2009. 2009.

[15] Qu, Bingqing, et al. "Content-based unsupervised segmentation of recurrent TV programs using grammatical inference." Multimedia Tools and Applications 76.21 (2017): 22569-22597.

[16] Wu, Xiaomeng, and Shin'ichi Satoh. "Ultrahigh-speed tv commercial detection, extraction, and matching." IEEE Transactions on Circuits and Systems for Video Technology23.6 (2013): 1054-1069.