

TV Ad Detection Using the Base64 Encoding Technique

Waseemullah

Department of Computer Science and IT
NED University of Engineering & Technology
Karachi, Pakistan
waseemu@neduet.edu.pk

Maria Andleeb Siddiqui

Department of Software Engineering
NED University of Engineering & Technology
Karachi, Pakistan
mandleeb@cloud.neduet.edu.pk

Muhammad Faraz Hyder

Department of Software Engineering
NED University of Engineering & Technology
Karachi, Pakistan
farazh@neduet.edu.pk

Muhammad Mukarram

Department of Computer Science and IT
NED University of Engineering & Technology
Karachi, Pakistan
mukki0303@gmail.com

Abstract-Automatic TV ad detection is a challenging task in computer vision. Manual ad detection is considered a tedious job. Detecting advertisements automatically saves time and human effort. In this paper, a method is proposed for detecting repeated video segments automatically, since generally, ads appear in TV transmissions frequently. At first, the user is allowed to browse the advertisements needed to be detected, and the video in which they are to be detected. The videos are then converted into a text file using the Base64 encodings. In the third step, the advertisements are detected using string comparison methods. In the end, a report, with the names of the advertisements is shown against the total time and the number of times these advertisements appeared in the stream. The implementation was carried out in python.

Keywords-TV ads; ad detection; base64

I. INTRODUCTION

Advertisements displayed in TV broadcasts are a very important part of a transmission as most of the revenue of a broadcaster is generated by advertising. Fast and accurate advertisement discovery is an important issue in the computer vision field. The main challenge of ad detection is the lack of information about the TV transmission structure and the unpredictable appearance of advertisements in the transmissions. In this paper, a novel approach is developed for TV ad detection. An algorithm has been developed for TV commercial (ad) detection. The proposed algorithm breaks the videos into frames. The sub-regions of frames take part in the comparison. The pixel values of the middle region of the frames of the TV transmitted video and the ad segment are compared. This approach results in TV ad detection with a 60% precision. Meanwhile, it suffers from the problem that the results rely heavily on the size of the video file. Another factor that significantly influences ad detection is the quality of the video. The proposed algorithm does not provide good results on grainy and poor-quality videos. The developed system based

on the novel approach for TV ad detection can be used by the PEMRA (Pakistan Electronic Media Regulatory Authority) to identify particular ads and their statistics. The framework is also usable for advertisement's agencies to compute and visualize the air time used for different ads. The experimental results confirm the importance of the proposed framework. This study may be helpful to (Figure 1):

- TV ad detection and identification: Many companies in marketing research and advertisement are keen on identifying commercial segments from TV broadcasts to verify the number of times a particular ad has been aired.
- TV advertiser concerns: A product owner has to keep a close eye on their competitors' marketing tactics by knowing the number of advertisements being aired per day and the total amount spent for advertising the products of a particular vendor.
- TV consumer woes: To give viewers an uninterrupted service is an expensive luxury nowadays as the bombardment of ads in every program is taking a toll on audience viewing. A viewer may find it easier to avoid any such occurrence in advance if the system is equipped with a consumer-friendly tool that reduces the ads' nuisance in the future.
- Market analysts inspection: In relation to the revenue aspect, a media monitoring agency such as PEMRA may need to know the broadcasted time ratio for ad and non-ad transmission of a TV channel. The resulting information may help the agency answer questions such as how much revenue a TV channel is generating through advertisements and which advertiser is paying higher for advertisements.
- Government regulatory body surveillance: Government agencies have to ensure that there should be no breaching in regulation of enforced laws during channel transmission.

Corresponding author: Waseemullah

One such agency would find useful a video segment identification application. On the other hand, many media observatory and regulatory bodies, want to know the real revenue generated by the various media channels.



Fig. 1. The possible applications for TV commercials detection.

Many approaches based on different features have been discussed [1, 2] for video detection and classification tasks. Authors in [3] investigated the problem of retrieving ads depending on their salient semantics. Semiotic is the knowledge of signs, associating signs with their meaning by communal conventions and particular cultural context. Identifying cuts, dissolve, and rhythm is used for video segmentation. Hough transform is used for calculating significant line slopes for describing shot content cluster analysis. Practical, playful, topic, and critical were recognized as semiotic categories of commercials. Evaluation was conducted on 150 ads from different Italian channels in compliance with the one conducted by human experts. The system showed best results on playful and worst results on practical ads. Authors in [4] explored the problem of real-time commercial detection using MPEG features in MPEG compressed videos. Black frame, unicolor frame, and change in aspect ratio were used for ad detection. It was shown that the least duration for any ad is one minute. Also, it was found that the strongest ad detecting parameter is the presence of black frames in commercial breaks. Authors in [5] proposed a new learning-based approach for ad detection. The approach uses several audio and visual features for classification based on Support Vector Machine (SVM). Average of edge change ratio, the variance of edge change ratio, average of frame difference, and variance of frame difference were used as visual features. NBC, ESPN2, and CNN TV channel transmissions of 10.75 hours was recorded for the evaluation, including different genres, i.e. movies, sports, and news. Recall values of 88.21% and 91.77% were observed without and with post-processing respectively. The respective precision values were 89.39% and 91.65%. Authors in [6] investigated the problem of identification and categorization of ads from TV transmission. Boundary detection was performed through a multi-modal approach. The ads were separated through black frame and silence features. Classification was conducted through text

detection. In addition, the absence of a channel logo was also used for ad segmentation from the normal transmission. Hidden Markov model based on audiovisual features was used for training. The results showed a precision value of 90% and a recall value of 80%.

Authors in [7] proposed a commercial detection that is depending on cookery/cooking programs. Audio-visual features were used for commercial boundary detection, including zero crossing rate, short time energy, edge detection, and corner detection. Furthermore, the logo of the program name was matched with a commercial break. Authors in [8] proposed a signal-based approach that uses an automatic unsupervised method for the segmentation of TV transmission. Using the general likelihood ratio and the Bayesian information criterion, this approach can be applied to audio signals, visual signals, or their combination. The evaluation was performed on recordings from French TV and the TRECVID dataset with recall values of 93% and 89% respectively. The respective precision was 93% and 91%. An approach based on Optical Character Recognition (OCR) and Automatic Speech Recognition (ASR) techniques for the classification of TV commercials was proposed in [9]. The output transcripts generated by these technologies normally produce a small number of keywords. These keywords are used to perform searching which results in semantically relevant information from the web. After that, retrieved information from the web was used to develop a feature vector based on text, and the commercial classification was performed based on text. The experimental results show that combining external resources improved the classification accuracy and helped avoiding speech recognition issues in TV commercial videos. A methodology based on subtitle detection to detect TV commercials in videos was presented in [10]. Many constraints were used to distinguish subtitles and other text that appeared in a frame. After detecting subtitles, a scheme was presented that decides whether a TV commercial exists or not, based on the appearance of subtitles. Then, the genetic algorithm was used to point out the mark-in and mark-out points. The reported precision and recall values were more than 90%. Authors in [11] evaluated an algorithm for the detection of text that normally cannot be detected easily in a video. This algorithm performs text detection in cases like a scene with high texture background. The reported results were 96% and 82% for precision and recall respectively.

Authors in [12] introduced a framework for the automatic semantic annotation of unconstrained videos. The framework helps to minimize the semantic gap, i.e. the difference between the low-level visual information and the corresponding human perception. It was also proposed that integrating visual similarity matching with common sense semantic relationships is a highly effective approach to automated video annotation. Authors in [13] investigated the problem of automatic management of videos taking into account syntactic and semantic features for TV NEWS programs. Authors in [14] presented an innovative solution of a smart emotional system for impaired people. They aimed to accompany the cognitive information contained in a movie, with the affective content for emotion recognition. The author in [15] described the TV stream as a collection of programs (P) and breaks

(advertisements) (B) [15]. A TV stream is a collection of program segments and break segments where C.1, C.2, and C.3 represent the different type of program segments having an aim of information or entertainment (Figure 2 in [15]). Each category of program segments is separated by breaks represented through C.

II. METHODOLOGY

In this research work, Base64 Encoding and Decoding technique has been used to detect TV advertisements. The following section defines the technique first and the steps carried out to perform the advertisement detection.

A. Base64 Encoding and Decoding

It is an encoding style that uses 64 numbers of characters to represent the binary data. Initially, data were sent over the network, i.e. via email through only a set of text characters but after the advancement in multimedia technology, it was expected to send other data types, i.e. images (binary data), attachments, and executables as well. Thus the need emerged for some encoding techniques for binary data to transport over the network. The encoding of binary data is needed to avoid the problem caused by the existence of null characters in binary data. The encoding Table of Base64 is shown in Figure 2.

Value	Char	Value	Char	Value	Char	Value	Char
0	A	16	Q	32	g	48	w
1	B	17	R	33	h	49	x
2	C	18	S	34	i	50	y
3	D	19	T	35	j	51	z
4	E	20	U	36	k	52	0
5	F	21	V	37	l	53	1
6	G	22	W	38	m	54	2
7	H	23	X	39	n	55	3
8	I	24	Y	40	o	56	4
9	J	25	Z	41	p	57	5
10	K	26	a	42	q	58	6
11	L	27	b	43	r	59	7
12	M	28	c	44	s	60	8
13	N	29	d	45	t	61	9
14	O	30	e	46	u	62	+
15	P	31	f	47	v	63	/

Fig. 2. Base64 encoding and decoding values.

B. The Followed Approach for TV Advertisement Detection.

The first technique that we tried was to compare and count the times the frames of an ad appear in the main broadcast video. Drawbacks such as high storage utilization, due to the splitting of videos into frames and more time, occurred. The second technique that we tried to use was to detect the advertisements during a live broadcast. The main problem that occurred was more time utilization although less storage was needed. The third technique was to encode all the pixels of all the frames, of all the videos, and then perform the comparison. This required more storage utilization as the size of the produced file was large, and more time in performing the comparison. The final approach chosen for this project was to encode only selected pixels belonging to sub-regions of the frames of the videos and then perform the comparison in order to reduce the computation cost. The time and the storage capacity required in this approach were comparatively lesser than the approaches mentioned above. The proposed methodology is shown in Figure 3. In Figure 3, the "Main File" is the video file that contains both advertisements and other program segments whereas the "Ad File" denotes the separate categories of advertisements that may exist in the input video

file. Both advertisement files and the Main File are converted into encoded text files using the Base64 encoding. The resulting text files are compared line by line to find the match of any advertisement file in the Main File. Each line of the text file corresponds to a set of pixel positions in a frame. These pixel positions are picked from the middle region of the frame only.

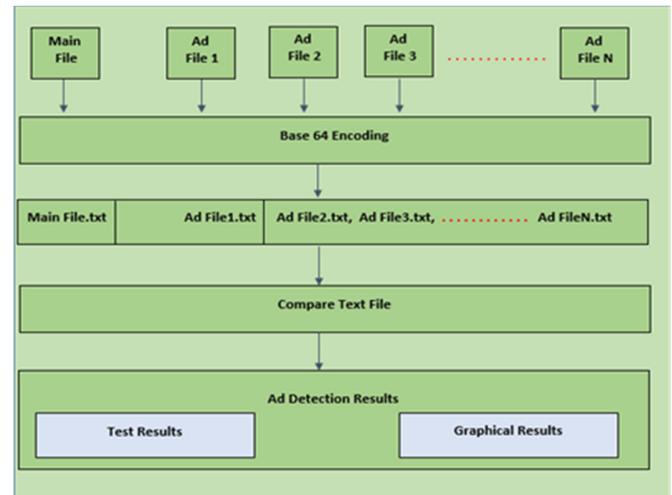


Fig. 3. The proposed approach for TV advertisement detection.

The obtained results were further presented in both textual and graphical forms. When the user selects the advertisements to be detected and the video they are to be detected in, instead of encoding each pixel of every frame, only selected pixels are encoded and saved into a text file and the file size remains small even for long videos. When the user hits the detect button, each pixel in each advertisement file is compared to each pixel in the Main video File until a match is found. When each pixel in the advertisement file is found in the main video file in sequence, the count of the comparison for that specific advertisement increases. Once the comparison is done for all selected ads, a table consisting of the ads against the time (seconds) they have appeared in the main video is shown to the user, with a graphical donut chart representation of the obtained results. Details are given in the result section of this paper. The steps of the algorithm are:

Step 01: Break the input videos (Main video and ad videos) into frames.

Step 02: Select the set of pixel values from each frame belonging to the sub-region of the selected frames.

Step 03: Convert the set of pixel values into base64 type.

Step 04: Compare all pixel values of an advertisement frame with the frame of the Main video

If the pixel points match then

Take the frame as an Ad frame and move to the next frame

Repeat the process until the Ad frame matches with the frames of the Main Video

Return the Total Number of Frames matched

Else

Select the next frame in the sequence

C. End-User System

A system has been developed for the end-user to evaluate the performance of the devised algorithm. The developed system is mainly comprised of four steps.

Step 01: In this step, the end-user is provided with the Ad Detection screen. The Select Main File button allows the user to select the video file which contains advertisements and non-advertisement content. The Select Ad Files button allows the user to either select a single advertisement segment or set of advertisement videos to be searched in the Main File (Figure 4).

Step 02: In this step, after providing the Main File and the Ad file, the end-user is provided with the Upload button. The button converts the videos into frames and then converts these frames into Base64 encoded strings (Figure 5).

Step 03: In this step each file is converted into a Base64 encoded text string (Figure 6).

Step 04: The screen in Figure 7 appears when the binary images are successfully converted into text files using the Base64 encoding. The end-user is provided with the Detect button to start the advertisement detection process with the developed algorithm. The system will produce two different forms of detection results, one in text form and another in graphical form.

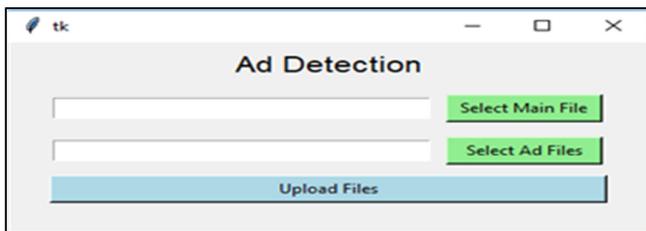


Fig. 4. Main screen of the system.



Fig. 5. Selecting the main video file and the advertisement file.

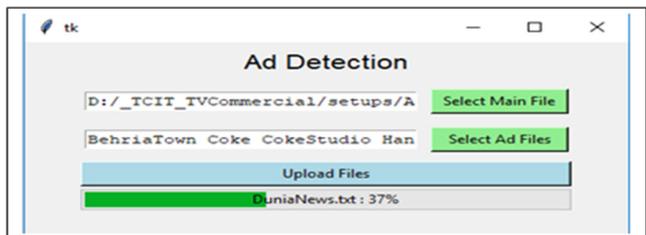


Fig. 6. Converting the main video file and the advertisement file into text files.

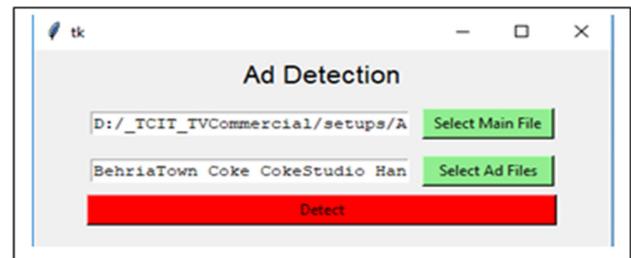


Fig. 7. After the successful conversion, the Detect button appears.

D. Dataset

The Video dataset was recorded from National TV channel transmissions, including ARY Digital 1, and Hum TV. The test data composed of three segments of transmission that were recorded from ARY had 2.5 hours length and the HUM TV segment had a length of 1 hour 40 minutes. All three segments were recorded from a local cable provider and were sampled at 25 frames per second.

III. RESULTS AND DISCUSSION

The detection results were obtained in two ways: single ad and multiple ad detection as shown in Figures 8-11. Figure 8 shows the text file output of the results and Figure 9 the produced graphical representation. They shows the detection of a KFC advertisement which lasts for 10 seconds in a given 152 second video file.

Detected Ads		
Ads Collection	Duration (sec)	
0	KFC	10

Fig. 8. Single ad detection text results.

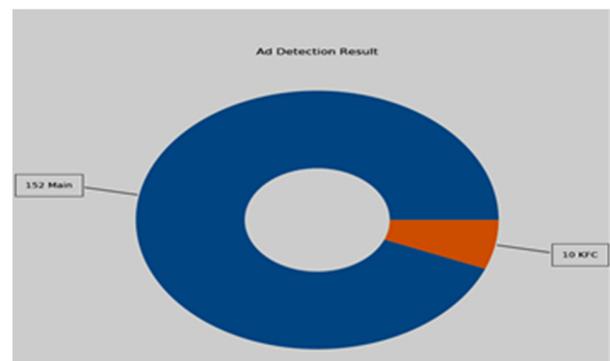


Fig. 9. Single ad detection graphical results.

Figure 10 shows the text file output of the results of multiple ad detection and Figure 12 their graphical representation. They show the number of detected advertisements. Moreover, the name of each detected advertisement is given in the adjacent column along with their duration in seconds in another adjacent column. To the best of our knowledge, this is the first time the Base64 encoding technique is used for TV advertisements detection. The

technique results in TV advertisement detection with 60% precision. On the other hand, it suffers from the problem that the results rely heavily on the size of the video file. One more factor that influences advertisement detection is video quality. The proposed algorithm does not provide good ad detection results on grainy and poor-quality videos.

Detected Ads		
Ads Collection	Duration (sec)	
0	KFC	10
1	Twisto	13
2	Zameen	15
3	Tapal	6
4	Coke	13
5	HanifJewel	5
6	BehriaTown	173
7	CokeStudio	343

Fig. 10. Multiple ad detection text results.

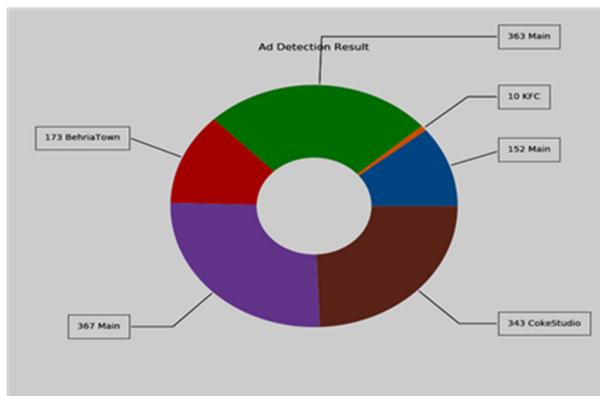


Fig. 11. Multiple ad detection graphical results.

IV. CONCLUSION

Many attempts have been made to find suitable features for TV commercial detection. It had been observed that some features that have been used by other researchers, such as silence, blank frame, and absence of TV channel logo during advertisement breaks have no use in Pakistani TV transmissions. It had been noticed that silence may appear not only at the beginning and the end of an advertisement but also at several intervals during the shot transitions in an advertisement. The blank frame that exists in TV transmission in Australia and European countries has been used by some researchers to detect the advertisement breaks, but Pakistani TV transmission does not contain such information due to the lack of media industry legislation in Pakistan. Another technique that is widely used in American TV channels is to detect the TV advertisement by finding the absence of the TV channel logo during the advertisements. Also, this is not true in Pakistani TV channel transmission, since the channel logo remains during the advertisement. In this paper, TV advertisement detection with Base64 encoding of the video segments was used. This saves computations as the ad detection is performed on text rather than on each pixel of an image. It also saves memory as only a very small set of pixels is used each time. The results can be improved by further reducing the sub-regions of images while converting them into text files. This can save memory and further reduce the

computation cost but it also causes the loss of important information needed for ad detection.

REFERENCES

- [1] S. Sahel, M. Alshafi, M. Alghamdi, and T. Alsubait, "Logo Detection Using Deep Learning with Pretrained CNN Models," *Engineering, Technology & Applied Science Research*, vol. 11, no. 1, pp. 6724–6729, Feb. 2021, <https://doi.org/10.48084/etasr.3919>.
- [2] P. Matlani and M. Shrivastava, "An Efficient Algorithm Proposed For Smoke Detection in Video Using Hybrid Feature Selection Techniques," *Engineering, Technology & Applied Science Research*, vol. 9, no. 2, pp. 3939–3944, Apr. 2019, <https://doi.org/10.48084/etasr.2571>.
- [3] C. Colombo, A. D. Bimbo, and P. Pala, "Retrieval of Commercials by Semantic Content: The Semiotic Perspective," *Multimedia Tools and Applications*, vol. 13, no. 1, pp. 93–118, Jan. 2001, <https://doi.org/10.1023/A:1009681324605>.
- [4] N. Dimitrova *et al.*, "Real time commercial detection using MPEG features," in *Proceedings of the 9th International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems (IPMU2002)*, 2002.
- [5] X.-S. Hua, L. Lu, and H.-J. Zhang, "Robust learning-based TV commercial detection," in *2005 IEEE International Conference on Multimedia and Expo, Amsterdam*, Netherlands, Jul. 2005, <https://doi.org/10.1109/ICME.2005.1521382>.
- [6] L.-Y. Duan, J. Wang, Y. Zheng, J. S. Jin, H. Lu, and C. Xu, "Segmentation, categorization, and identification of commercial clips from TV streams using multimodal analysis," in *Proceedings of the 14th ACM international conference on Multimedia*, New York, NY, USA, Oct. 2006, pp. 201–210, <https://doi.org/10.1145/1180639.1180697>.
- [7] N. Venkatesh, B. Rajeev, and M. G. Chandra, "Novel TV Commercial Detection in Cookery Program Videos - PDF Free Download," in *Proceedings of the World Congress on Engineering and Computer Science 2009*, San Francisco, CA, USA, Oct. 2009, vol. 2.
- [8] E. El-Khoury, C. Sénac, and P. Joly, "Unsupervised Segmentation Methods of TV Contents," *International Journal of Digital Multimedia Broadcasting*, vol. 2010, Jun. 2010, Art. no. e539796, <https://doi.org/10.1155/2010/539796>.
- [9] Y. Zheng, L. Duan, Q. Tian, and J. S. Jin, "TV Commercial Classification by using Multi-Modal Textual Information," in *2006 IEEE International Conference on Multimedia and Expo*, Toronto, Canada, Jul. 2006, pp. 497–500, <https://doi.org/10.1109/ICME.2006.262434>.
- [10] Y.-P. Huang, L.-W. Hsu, and F.-E. Sandnes, "An Intelligent Subtitle Detection Model for Locating Television Commercials," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 37, no. 2, pp. 485–492, Apr. 2007, <https://doi.org/10.1109/TSMCB.2006.883428>.
- [11] L. Meng, Y. Cai, M. Wang, and Y. Li, "TV Commercial Detection Based on Shot Change and Text Extraction," in *2009 2nd International Congress on Image and Signal Processing*, Tianjin, China, Oct. 2009, <https://doi.org/10.1109/CISP.2009.5302320>.
- [12] A. Altadmri and A. Ahmed, "A framework for automatic semantic video annotation," *Multimedia Tools and Applications*, vol. 72, no. 2, pp. 1167–1191, Sep. 2014, <https://doi.org/10.1007/s11042-013-1363-6>.
- [13] J. Wang, M. Xu, H. Lu, and I. Burnett, "ActiveAd: A novel framework of linking ad videos to online products," *Neurocomputing*, vol. 185, pp. 82–92, Apr. 2016, <https://doi.org/10.1016/j.neucom.2015.12.038>.
- [14] D. Affi, J. Dumoulin, M. Bertini, E. Mugellini, O. Abou Khaled, and A. Del Bimbo, "SensiTV: Smart EmotionNal System for Impaired People's TV," in *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video*, New York, NY, USA, Jun. 2015, pp. 125–130, <https://doi.org/10.1145/2745197.2755512>.
- [15] Z. A. A. Ibrahim, "TV Stream Table of Content: A New Level in the Hierarchical Video Representation," *Journal of Computer Sciences and Applications*, vol. 7, no. 1, pp. 1–9, Dec. 2018, <https://doi.org/10.12691/jcsa-7-1-1>.