IRAQI Academic Scientific Journals

Available at https://www.iasj.net/iasj
Iraqi Academic Scientific Journals
Journal homepage:
https://journals.uokerbala.edu.iq/index.php/UOKJ

Research Article

A Proposed Recommendation System to Exhibit Product Advertising at Proper Time Stamps of Video

[1,]Heba Adnan Raheem [2,] Tawfiq A. al-assadi
[1,] Computer Science Department, Collage of Computer Science and Information Technology University of Kerbala, Karbala, Iraq
[2,]software Department, College of Information Technology, Babylon University , Babil,Iraq.

Article Info
Article history:
Received 12 -1-2025
Received in revised form 6 -2-2025
Accepted 19-2-2025
Available online 13 -4 -2025
Keywords : Feature extraction, Online videos, Product advertising, Recommendation system, Time stamps.

Abstract:

An advertisement is a notice or announcement in a public medium promoting a product. Advertising plays significant role in the introduction of a new product in the market. It stimulates the people to purchase the product and provides an opportunity for e-commerce companies to recommend their products in videos. In this research, we propose a video advertising system to exhibit appropriate video product ads to particular users at proper time stamps. This takes into account video content (semantics or relation between video and products), and viewing behavior (user and video). In the proposed framework, the first stage will be analyzed video structure into its primary components. Next stage will be associations between users and videos and videos and products. Finally, the recommendation system is proposed to integrate key frame-product association and the important shot of original video to recommend appropriate video of product advertisement(ads) to user. The results of determining the video category by employing the Yolo deep learning model (Yolov5s) in precision. Thus, the proposed method is more accurate in getting the video content. The results of the recommendation and video advertising algorithm demonstrate that all the recommended videos are accurately classified and suggested to the actual type of viewing videos in the GWO algorithm.

Corresponding Author E-mail : hiba.adnan@uokerbala.edu.iq , Tawfiqasadi@itnet.uobabylon.edu.iq
Peer review under responsibility of Iraqi Academic Scientific Journal and University of Kerbala.

1. Introduction

A tremendous amount of multimedia information including digital video is becoming prevailing as a result of the advances in multimedia computing technologies and high-speed networks.

The video can be defined as "a huge volume data object, it contains high redundancy and intensive information", it has a complex structure that consists of scenes, shots, and frames. Figure 1 shows the structure of the video.

Video e-Commerce, is presented for large scale online video advertising, which is able to exhibit appropriate product ads to particular users. In order to effectively mine the relationship among users, videos and products.

In addition, E-commerce companies like Amazon.com and video content providers such as Netflix have made the recommended systems salient parts of their websites. However, few works pay attention to the online video advertising with the combination of recommendation and video content analysis. There are several problems in these systems:

- The first is that most video ads inserted in video streams are intrusive and boring without considering user's attention.
- The second is that traditional video ads are inserted in the middle of the program or in highlights of the sports video undermines the continuity of the video program.
- The third is textual keywords used in online video advertising is not enough for measuring the relevancy of rich content videos.

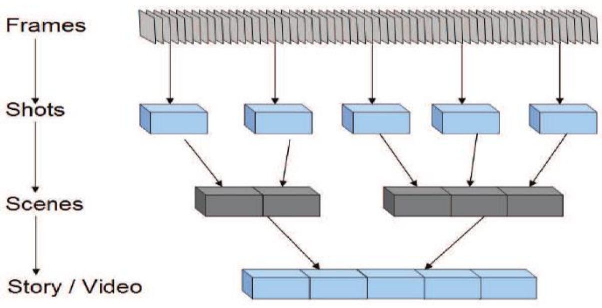

Figure 1: The structure of video [15]

There are different researches in this side are sorted as in [1] recommender systems (RSs) have been proposed to collect opinions and preferences about a set of items, process such preferences and build a personalized information access. They proposed framework is built on top of a Distributed Ledger technology platform that runs without any centralized authority and it supports both decentralized ratings and ranking of different items. A preliminary evaluation on the test network demonstrates the feasibility of the framework in terms of performance and cost.

In [2] they developed a method to jointly learn the index structure and user preference prediction model. They proposed joint optimization framework; the learning of index and user preference prediction model are carried out under a unified performance measure. Besides, we come up with a novel hierarchical user preference representation utilizing the tree index hierarchy. Experimental evaluations with two large-scale real-world datasets show that the proposed method improves recommendation accuracy significantly. Online A/B test results at a display advertising platform also demonstrate the effectiveness of the proposed method in production environments.

In [3] they introduced an approach to increase the accuracy and to improve the performance of collaborative filtering recommender system. A hybrid approach is proposed to improve the performance of video collaborative filtering recommender system based on clustering and evolutionary algorithm. Proposed approach combines k-means clustering algorithm and two different evolutionary algorithms which are Accelerated Particle Swarm Optimization Algorithm (APSO) and Forest Optimization Algorithm (FOA). The main aim of this paper is to increase the accuracy of recommendation of user-based collaborative filtering video recommender system. Evaluation and computational results on the Movie Lens dataset show that the proposed method has a better performance than the other related methods.

In [4] they adopted the grey wolf optimization (GWO) algorithm to select the optimal parameters of the proposed model, and they based on support vector machine regression model based on the grey wolf optimization algorithm (GWO-SVR). Three real bus lines were taken as examples to validate the model. The results show that the mean average percentage error is 14.47% and the mean average error is 0.7776. In addition, the estimation accuracy and training time of the proposed model are superior to the genetic algorithm-back propagation neural network model and grid-search support vector machine regression model.

In [5] they enhanced Grey Wolf Optimizer (GWO) for designing the evolving Convolutional Neural Network-Long Short-Term Memory (CNN-LSTM) networks for time series analysis. To overcome the probability of stagnation at local optima and a slow convergence rate of the classical GWO algorithm, the newly proposed variant incorporates four distinctive search mechanisms. The CNN-LSTM networks devised by the proposed (GWO) variant offer better representational capacities to not only capture the vital feature interactions, but also encapsulate the sophisticated dependencies in complex temporal contexts for undertaking time-series tasks.

In [6] they proposed a novel advertising strategy for the rec/ads trade-off. To be specific, they developed an RL-based framework that can continuously update its advertising strategies and maximize reward in the long run. Given a recommendation list, they designed a novel Deep Qnetwork architecture that can determine three internally related tasks jointly, i.e., (i) whether to interpolate an ad or not in the recommendation list, and if yes, (ii) the optimal ad and (iii) the optimal location to interpolate. The experimental results based on real-world data demonstrate the effectiveness of the proposed framework.

In [7] they described Video e-commerce, a method for online video advertising that is able to display relevant product adverts to specific consumers based on video content. In order to accurately model the connections between people, videos, and goods, the authors suggest using Co-Relation Regression, a Heterogeneous Information Network, and Video Scene Importance. This guarantees product variety, solves data sparsity and cold-start issues, and highlights the significance of the video. Finally, product recommendations are made using Heterogeneous Relation Matrix Factorization. Currently, this method is being tested on a publicly available dataset. It aims to propose product adverts to specific clients at optimal times. However, the speed of the complete framework may be improved by investigating the potential of large-scale web advertising and performing real-time updates for new items.

In [8] they expanded their aforementioned previous work in [7] using an iterative Co-Relation Regression (ICRR) model that can efficiently mine the associations between users, videos, and goods. To accommodate the need for massive amounts of internet advertising, the model effectively adjusts to new key frames or new things that decrease the complexity of time. In this way, the whole architecture can be tested in real time using data available online. One potential drawback of this work is the loss of personal privacy that may arise from a viral film being widely distributed online. In my perspective, secure communication techniques are necessary for this project.

In this paper, improving the proposing a recommendation system to exhibit product advertising at proper time stamps of video has been proposed. It is totally presented as follows: Section 1 is introduction, Section 2 is the proposed method, Section 3 is method, Section 4 is results and discussion, and Section 5 is conclusion.

2.The Proposed Method

The proposed system is based on video advertising system to exhibit appropriate video product ads to particular users at proper time stamps of videos, which takes into account video content (semantics or relation between video & products), and viewing behavior (user & video). In the proposed framework, first state will be video structure analysis into its primary components. Next state will be the associations between users and videos, videos and products. Lastly, a recommendation system is proposed to integrate key frame-product association, and video shot importance to recommend appropriate video of product ads to users. Figure 2 showed the general overview of the proposed system:

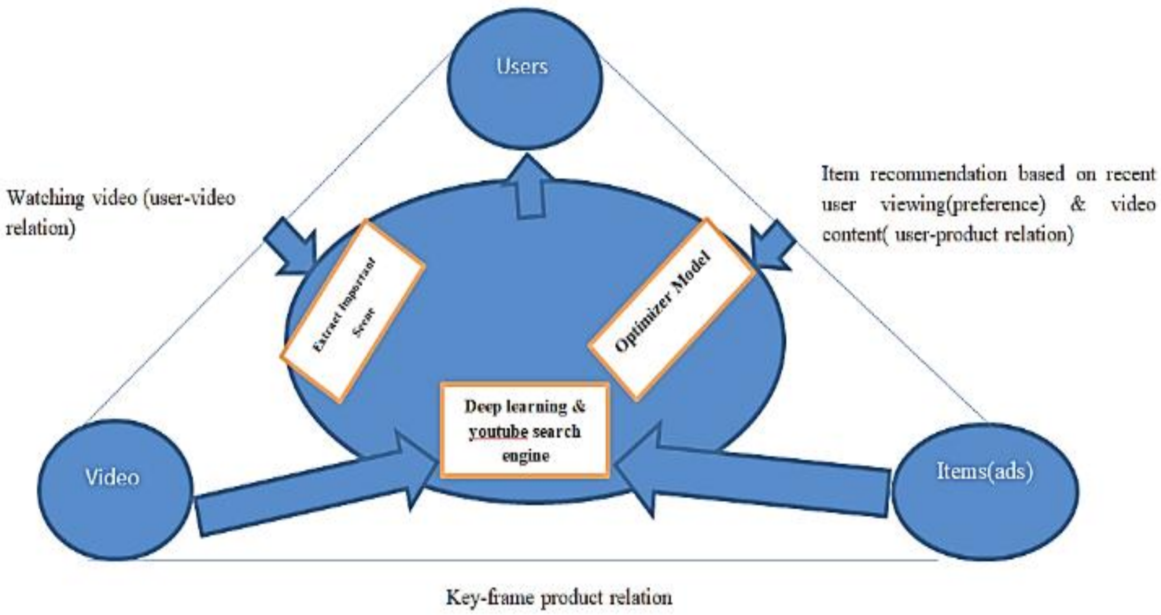

Figure 2: The general overview of the proposed system

As shown in Figure 3, the proposed system is implemented based on following main stages as follow:

1. Important Shots and Video Summarization in order to find a proper time stamps of videos, and represent the associations between users and original videos.
2. Class Detection and online Deep Learning model to obtain the content of the original video with YouTube search engine to find the relation between videos and products which are considered the first stages of recommended process.
3. Building an online video advertising system by recommendation.

It is proposed to integrate key frame-product association, and video shot importance to recommend appropriate product ADS to users. We faced several challenges and obstacles while implementing this system, including:

1. Video content analysis and get high accuracy for object detection using yolo deep learning.
2. The method of collecting information to determine the type of video through the frames, in other words, determining the type or content of the video through the objects in each frame.
3. Dealing with YouTube, as it always gets updates that cause a problem in dealing due to YouTube policies.
4. Obtaining a dataset or videos that represent the views of the user so that it contains multiple topics and not a specific topic, and this was the biggest challenge because most sites contain videos for a specific topic. Also, advertising videos do not have

free links to download them because they have a financial cost.

5. The proposed system is implemented with Python programming language and YouTube API application So, the proposed system methods consist of the following implementation steps:

- Step 1: Using the YouTube data API v 3.0 to connect python with YouTube servers.
- Step 2: Select the original video using GUI and analyze the video into its important components.
- Step 3: Extract interested points (features) from each frame in the shot for an original video by using ORB algorithm and compute the similarity between features using FLANN, then summary using the entropy has been done to find the important shot and its key-frame in the video.
- Step 4: object detection to obtain the content of the original video based on yolo deep learning model (Yolov5s) from https://github.com/ultralytics/yolov5/releases. It is clearly shown that the online deep learning model consists of many predefined models. it can predict the class type for the uploaded video.
- Step 5: Searching the YouTube channel about the advertisement video to bring the ads videos similar to uploaded video.
- Step 6: YouTube search engine will reply with list of videos. We analyze the YouTube search results and compare between the time of the suggested advertisement video and the main video time.
- Step 7: In order to do the recommendation process, the used model entered the list of advertising videos that have been retrieved from YouTube channel on the recommendation algorithm based on an optimizer model for the purpose of returning the appropriate advertising video to the original video. You Tube search engine use the user Google account to optimize the search result, this information collected using the YouTube data API. In addition, other parameters can consider (the number of likes, ID of video, number of watching, and no. of comments).
- Step 8: Play the main video and embed the advertisement video before the importance shot.
- Step 9: Play the recommended advertisement video then play the rest of the main video.

Figure 3 shows the stages which demonstrates the methodology of the proposed framework which it is consist of summarization part and an online video advertising system by recommendation part.

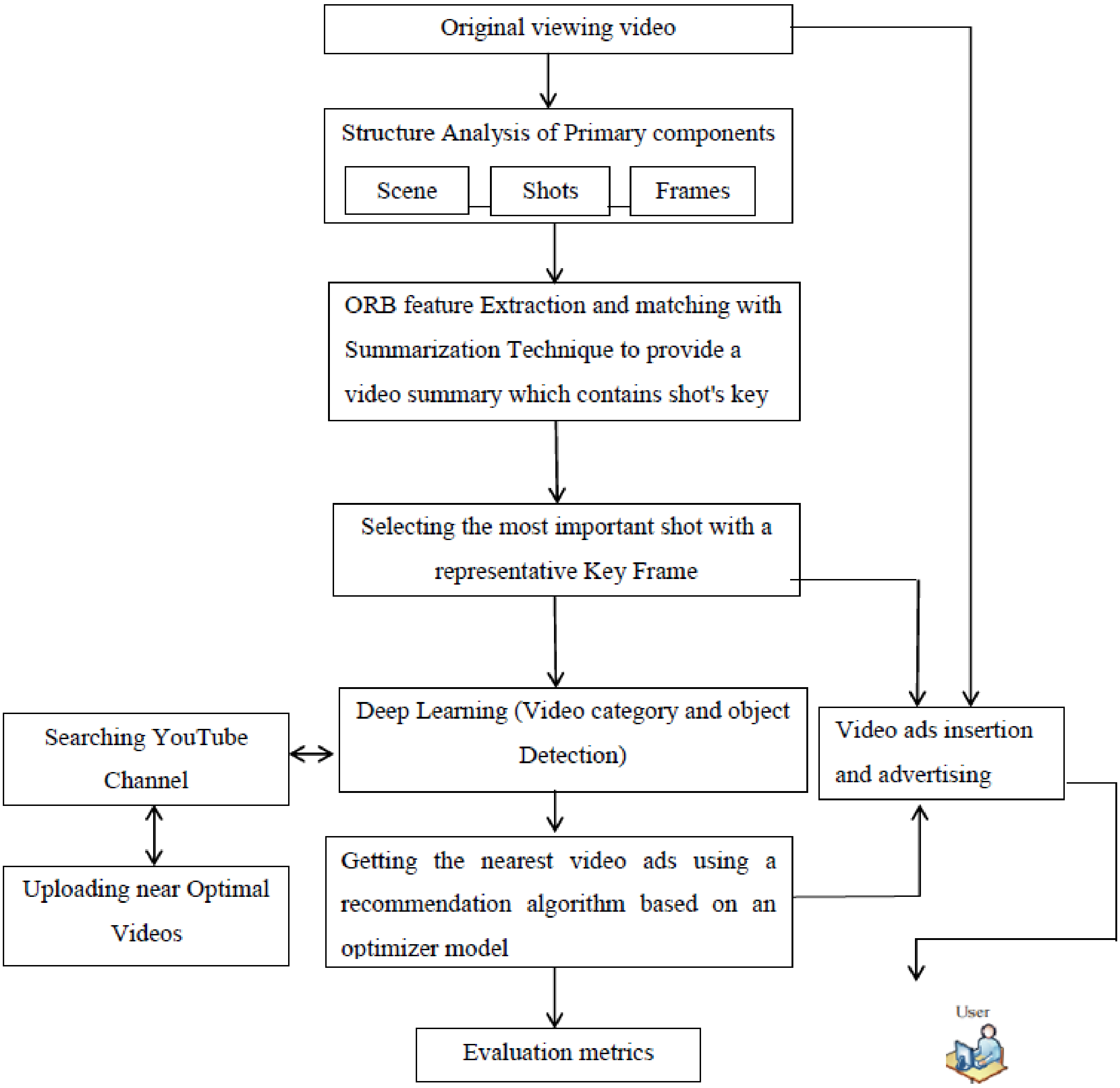

Figure 3: The proposed system stages

2.1. Data set description

In this part, the datasets that were utilized to test the suggested system are described. We employed the widely used TVSum50 Benchmark Dataset [9] in the suggested research for video summarization and shot extraction. This dataset contains 50 video sequences collected from YouTube using 10 categories from the TRECVid MED task. These videos cover a wide range of topics, including flash mobs, news, video blogs, and more. A self-built video dataset that was gathered from the internet was used to verify the method's viability. We utilized the SumMe [10] dataset for additional testing, which was constructed as a standard for video summarizing. The 25 videos in the collection cover sports, events, holidays, and other topics. The videos range in length from 1 to 6 minutes.

Further, a collected video dataset that was gathered from the internet was used to verify the method's viability. The popular online video collections were utilized for getting advertisements and YouTube videos dataset from the most popular video sharing site.

The proposed system methodology consists of Oriented Rotated Brief Very Fast Binary Descriptor (ORB) algorithm for feature extraction and matching techniques to important determine shots for all video. ORB based on Fast Library for Approximate Nearest Neighbors (FLANN) technique to classify frames into tree and comparison among features.

2.2.Analysis of the video structure into its primary components

This step involves splitting the video into its major components. The structure of the viewing video file will be analyzed in order to obtain its primary components (scenes, shots, frames) so as to apply the ORB method in the next step, and as the following steps:

- Convert: Video to sequence of frames.
- Transform: Frames (Bitmap color image (24 bits) to 8-bits (Gray scale) by YCbCr Color Format.

Color Layout Descriptor (CLD) is used to determine color spatial distribution. The frames are first taken from the video. The retrieved frames are color space transformed from RGB to YCbCr. Each YCbCr frame is divided into 8x8 non-overlapping blocks.

Gray scaling is the process of transforming an image (frame) from a color system such as Red, Green, and Blue (RGB) to shades of gray. RGB is the most common means of encoding a pixel in a color space. One of the motivations for choosing gray scale photographs rather than color images is to reduce the RGB color space to a single dimension description, i.e.

monochromatic. This is required to reduce calculation costs and simplify processing. Some applications do not require color images, and utilizing them may increase the number of training data required to get effective results.

The SBD technique attempts to divide the video into temporal parts (shots). This stage seeks to identify the sequence of frames acquired by a single camera operation. Frames in the same picture frequently feature significant contrast. correlation, and the SBD step determines the boundaries between shots.

2.3.Important shots and video summarization

There are two components to the video summarizing technique in our suggested study. Using shot boundary identification based on ORB characteristics, the entire video is first divided into separate video segments (scenes and shots). After that, a summary for each scene is applied such that each shot's key frame extraction is realized using the enhanced clustering technique based on entropy metric calculation. Finally, by organizing key frames sequentially, a summary of the video is produced. Notably, important frames in this study refer to the representative frames that can represent the framework of the video contents as shown in Figure 4.

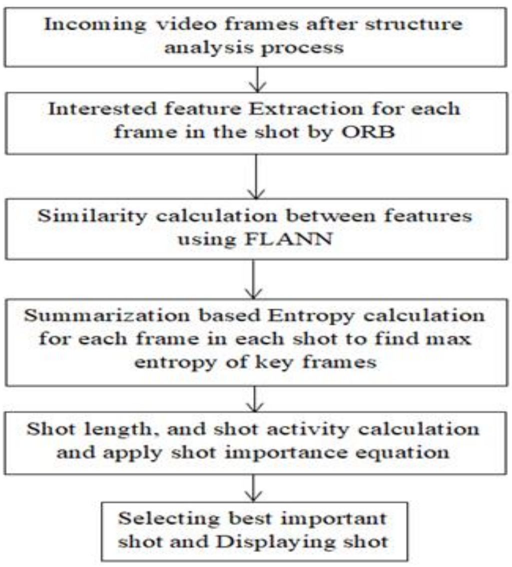

Figure 4: Framework of detecting the important shot based on video summarization

2.4.Deep learning and video class detection

The class and object detection need to be determined to obtain the content of the original video based on the yolo deep learning model (Yolov5s). It is clearly shown that the online deep learning model consists of many predefined models. To make the video analysis process more comprehensive, it is necessary to collect the information of each shot in a single 3D matrix. This matrix could be analyzed in many ways to obtain the highest possible accuracy to determine the category of the video by using a pre-trained model. The standard measures are the percentage of the object spread on the whole video. If the object is spread over the entire video, it is the real candidate and not the object being concentrated in one place. The frequency of appearance of this object within the video also needs to be taken into account. All of these metrics will help to determine the real category for the original video by analyzing its components as shown in Figure 5.

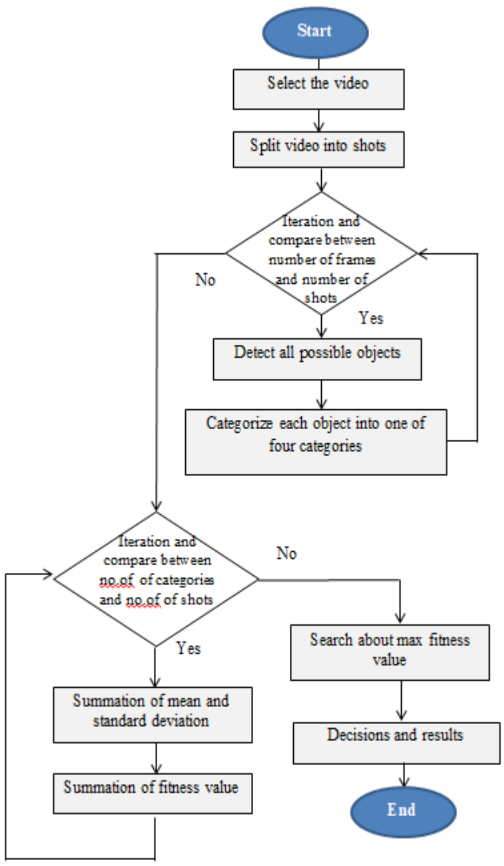

Figure 5: Video class detection algorithm based on yolo deep learning model

2.5. API YouTube channels and similar videos

After detecting the type of the entire video based on the video class detection in section 2.4, the YouTube channel is then searched for the advertisement video. The YouTube videos and its related metadata, i.e., user comments/opinions have become a treasure of the data. This data could be used in various fields of research especially in online video advertising.

The proposed framework consists of five steps as shown in Figure 6:

- Input parameters (video category, channel ID, and API key).
- Searching in the YouTube channel about the advertisement video.
- Collecting the similar ads videos.
- Video metadata collection.
- Replay list of videos with its meta data.

2.6. Recommendation based on grey wolf optimizer algorithm (GWO)

The concept of this algorithm depends on filtering or filtering the videos obtained through the two previous processes (object detection & you tube search engine) to find the closest ad video that is similar or appropriate to the original video content, in addition to having suitable parameters such as (likes, comments, index, and views). In other words, the (GWO) algorithm classifies advertising videos in a hierarchical manner, where the top of the pyramid is the most appropriate video.There are several stopping conditions for this algorithm as in following:

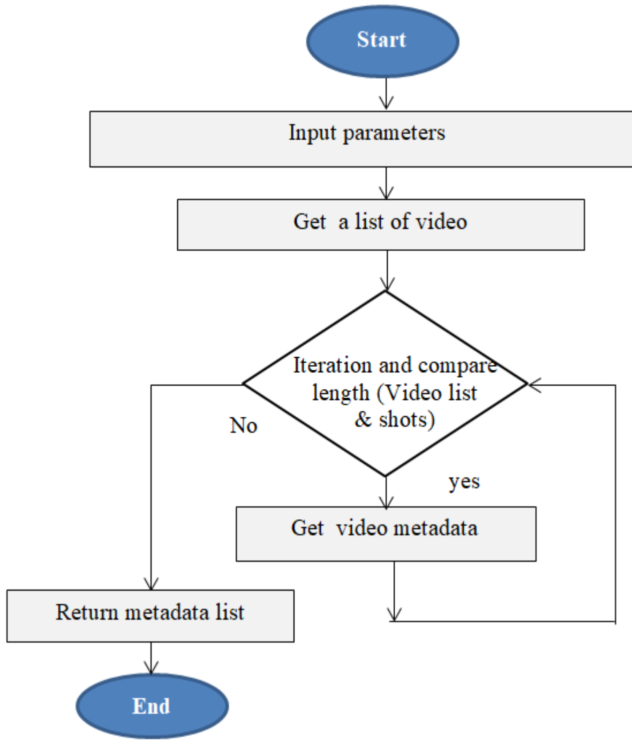

Figure 6: A frame work of collecting the similar ads videos and related meta data based on YouTube API

- Reaching a fixed number of iterations (i).
- Access to determined specifications such as (Fitness <= specific number).
- The difference between one iteration (i) and another is very small.

Figure 7 illustrates the framework of selecting the best ads ad for video using (GWO) algorithm.

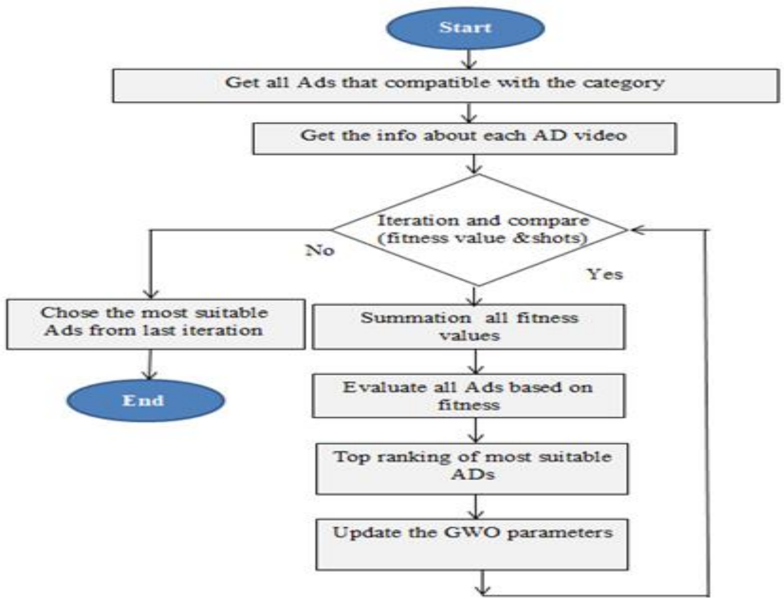

Figure 7: A framework of selecting the best ads ad for video using GWO algorithm

2.7. Online video advertising

Video Advertising and display appropriate product ads to users by embedding it at a proper Timestamp in the importance shot of video which it considered the final stage of the recommended process.

207. The first step is user-watching video relation, which indicates the commonalities between viewers and key frames. It is used to determine the best insert product adverts position using the summarizing approach.

208. The second one, known as key frame-product relation, is the relationship between key frames and products. It combines user preference with video semantics based on deep learning and YouTube search engine.

209. The third one is a recommendation for an item based on recent user watching and video content as shown in Figure 8.

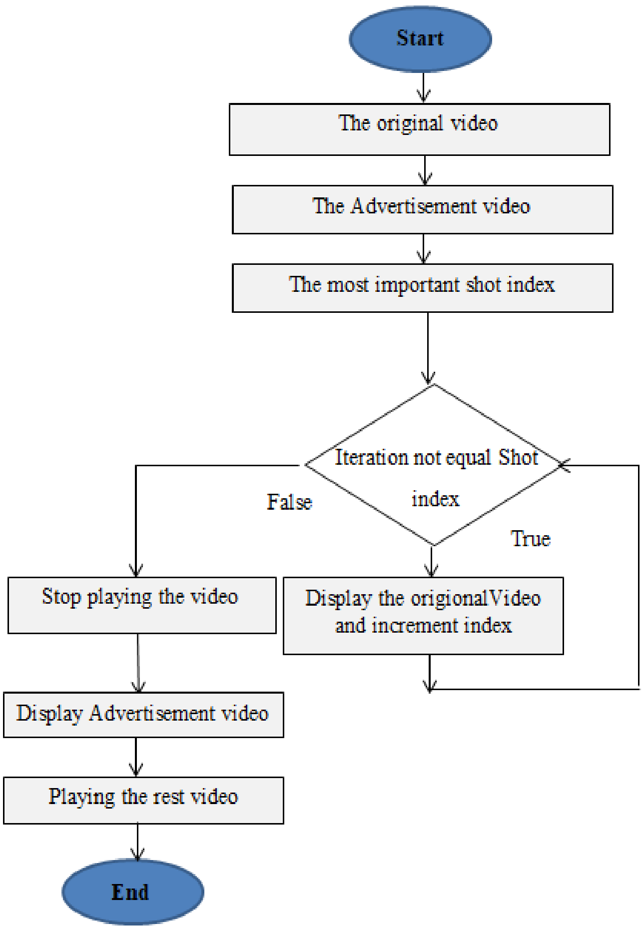

Figure 8: A frame work of an online video advertising display process

210. Results and Discussion

The proposed system of important shots based on video pre-processing, as it showed in Figure 9 after each frame of the video stream and the template image are first processed in gray scale as following:

Figure 9: Converting a video to sequence of true color image (frames)

In order to verify the processing efficiency of the algorithm, we calculated the running speed of video summarization in frames per second. The results are shown in Table 1. Table 2 compares the feature extraction speed with SIFT based algorithms by [11, 12].

Table 1: Running efficiency of our method

Detection of Shots		Extraction of Key Frames
Feature Extraction	Feature Matching	1045.18 frames/s
43.25 frames/s	30.62 frames/s	

Table 2: Comparison of running efficiency

Methods	Frames	Feature Extraction Speed
Proposed	1357	43.25 frames/s
[11]	1351	41.20 frames/s
[12]	1920	26.40 frames/s

It can be seen from the results that the speed of our algorithm in ORB feature point extraction is 43.25 frames per second, which faster than [11] and faster than [12], therefore, the ORB based algorithm has a speed advantage over SIFT based algorithm.

Recall and precision were used for evaluation metrics. Recall represents the ratio of the number of true positive shots to the total number of shots. Precision refers to the rate of the true shots that are detected.

In order to verify the effect of this algorithm on shot boundary detection, we tested the recall and precision under different types of videos, in comparison with the algorithms by [11-14] as shown in Table 3.

Table 3: Comparison of shot detection methods

Methods	Collected Dataset		Methods	TVSum50		SumMe	
proposed	Recall	Precision	proposed	Recall	Precision	Recall	Precision
	98.20%	95.30%		97.81%	90.87%	96.33%	93.20%
[12]	94.45%	95.79%					
[13]	96.40%	95.81%					
[14]	97%	95%					
[11]	97.20%	93.30%		95%	87%		

From Table 3, we can see that our method shows the advantages in recall rate on collected datasets, the TVSum50 dataset and SumMe dataset.

211.Experimental Results of The Recommendation and Video Advertising Algorithm

The evaluation measure (MSE and RMSE) for recommendation algorithm to suggest the appropriate recommender type of video advertisement for the actual type of viewing video is illustrated in Table 4.

Table 4: The Evaluation Measure for Recommendation Using (GWO) Algorithm

video no.	actual type	Recommended type	error	sqr error	mean sqrt error	root mean sqrt error
1	sporty	Sporty	0.0	0.0	0.0	0.0
2	sporty	Sporty	0.0	0.0		
3	sporty	Sporty	0.0	0.0		
4	sporty	Sporty	0.0	0.0		
5	sporty	Sporty	0.0	0.0		
6	sporty	Sporty	0.0	0.0		
7	trans	Trans	0	0	0..67	0.818
8	trans	Trans	0	0		
9	trans	Sporty	1	1		
10	trans	Trans	0	0		
11	trans	Trans	0	0		
12	trans	Trans	0	0		
13	cooking	Cooking	0	0	0	0
14	cooking	Cooking	0	0		
15	cooking	Cooking	0	0		
16	cooking	Cooking	0	0		
17	cooking	Cooking	0	0		
18	cooking	Cooking	0	0,67		
19	animal	animal	0	0	0	0
20	animal	animal	0	0		
21	animal	animal	0	0		
22	animal	animal	0	0		
23	animal	animal	0	0		
24	animal	animal	0	0		

As shown in table 4, we can see that all the recommended videos are accurately classified and suggested to the actual type of viewing videos. There is one error in recommended appropriate video ad for viewing video 9. On the other hand, when applying the same measures above to the comprehensive search algorithm, we have found that the error rate is completely zero, because it depends on searching and choosing between each ad and another in a sequential manner. So it is more appropriate with a few ad videos and takes more time complexity. This proves that the GWO results are realistic and more accurate.

212.Conclusion

The proposed system has more than one of main stages; firstly, video shot extraction based on video summarization in order to find a proper time stamps of videos. The second stage is to obtain the content of the original video base on deep learning with YouTube search engine to find the relation between videos and products. Finally, online video advertising system by recommendation is constructed. The results of the proposed system are:

213.ORB and FLANN local feature descriptors and matching can provide a good representation of the frame image. Despite the fact that these features have some limitations, their combination gives additional robustness for feature extraction. Also, combination (ORB) and (FLANN) with image enhancement improves feature extraction and matching to recognize the boundaries of shot or scene.

214.The suggested system is able to display suitable items (ads) to specific users at the appropriate time.

215.GWO results are realistic and more accurate in the case of the number of advertising videos being very large.

216.The suggested system is able to display suitable items (ad) to specific users at the appropriate time. For further development of the current research, the following suggestions are important to implement in the future:

217.Another method, such as graph mining, can be utilized to give a good video summary and detect the most important shot.

218.Exploring online advertisements at the large scale and performing real-time updates for new objects will enable the framework as a whole to perform effectively in real time.

219.Designing and implementing a security method in order to guarantee the user's privacy.

References

[1] Lisi, A., De Salve, A., Mori, P., & Ricci, L. (2019). A smart contract based recommender system. Economics of Grids, Clouds, Systems, and Services, Leeds, UK.

[2] Zhu, H., Chang, D., Xu, Z., Zhang, P., Li, X., He, J., Li, H., Xu, J., & Gai, K. (2019). Joint optimization of tree-based index and deep model for recommender systems. Advances in Neural Information Processing Systems.

[3] Tohidi, N., & Dadkhah, C. (2020). Improving the performance of video collaborative filtering recommender systems using optimization algorithm. *International Journal of Nonlinear Analysis and Applications*, 11(1), 483-495.

[4] Qin, W., Wang, L., Liu, Y., & Xu, C. (2021). Energy consumption estimation of the electric bus based on grey wolf optimization algorithm and support vector machine regression. *Sustainability*, 13(9), 4689.

[5] Xie, H., Zhang, L., & Lim, C. P. (2020). Evolving CNN-LSTM models for time series prediction using enhanced grey wolf optimizer. *IEEE Access*, 8, 161519-161541.

[6] Zhao, X., Gu, C., Zhang, H., Yang, X., Liu, X., Tang, J., & Liu, H. (2021). Dear: Deep reinforcement learning for online advertising impression in recommender systems. *Proceedings of the AAAI*

Conference on Artificial Intelligence, *35*(1), 750-758.

[7] Cheng, Z.-Q., Liu, Y., Wu, X., & Hua, X.-S. (2016). Video ecommerce: Towards online video advertising. Proceedings of the 24th ACM International Conference on Multimedia, Amsterdam, The Netherlands.

[8] Cheng, Z.-Q., Wu, X., Liu, Y., & Hua, X.-S. (2017). Video ecommerce++: Toward large scale online video advertising. *IEEE Transactions on Multimedia*, 19(6), 1170-1183.

[9] Song, Y., Vallmitjana, J., Stent, A., & Jaimes, A. (2015). TVSum: Summarizing web videos using titles. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

[10] Gygli, M., Grabner, H., Riemenschneider, H., & Van Gool, L. (2014). Creating summaries from user videos. Computer Vision–ECCV 2014, Zurich, Switzerland.

[11] Liang, B., Li, N., He, Z., Wang, Z., Fu, Y., & Lu, T. (2021). News video summarization combining surf and color histogram features. *Entropy*, 23(8), 982.

[12] Feng, L., Yan, Z., Shi-gang, W., & He-xin, C. (2016). Video scene mutation change detection combined with SIFT algorithm. *Chinese Optics*, 9(1), 74-80.

[13] Hannane, R., Elboushaki, A., Afdel, K., Naghabhushan, P., & Javed, M. (2016). An efficient method for video shot boundary detection and keyframe extraction using SIFT-point distribution histogram. *International Journal of Multimedia Information Retrieval*, 5, 89-104.

[14] Algur, S. P. (2016). Video key frame extraction using entropy value as global and local feature. *arXiv:1605.08857*.

[15] Y. Li and C. J. Kuo, Video content analysis using multimodal information: For movie content extraction, indexing and representation: Springer Science & Business Media, 2013.