Software Proposal Document for project InMovie Recommendation

Aly Mekawy, Amr Sherif, Foad Osama, Youssef Roshdy

October 2019

Abstract

Video streaming is becoming essential in most of users life , On most video streaming platform , users get their recommended videos based on some algorithms , calculations ,implicit feed-backs ,watch and search behaviors, Also new videos suffer from cold-start which happens to new videos when they just get uploaded as they have problem in which no data or reviews , so its harder to recommend to some-user, another problem users face everyday is we cant really find the desired content exactly unless labeled or has multiple views as the search engine will find the videos based on keywords or tags not really the content inside the video .

1 Introduction

1.1 Background

Generally , feature extraction involves analysing huge amount of data based on the objects detected from a frame of a video.generally early levels of feature extraction are algorithms like edge detection and thresholding all the way going up to computer vision , the overall goal of the feature is describing the in video data with great accuracy

1.2 Motivation

almost all of any video platform service uses algorithms related with numbers and calculations there is no really a way a user can find a related video with the actual content desired, market motivation is going for a video streaming platform or a video search engine

1.3 Problem Definitions

Enhancing the video recommendation system and improving the video classification precision with the ability to search specific scenes

2 Project Description

The system should able be detect the objects from the scenes of the video that a user inputs directly or its automatically chosen based on behavior of the user, then the objects in the scene classified and labeled, each object with its own label, from these data gathered, by using algorithms we classify the current scene based on the objects the system detected. The next phase uses feature extraction which each object will be ranked and based on these rankings an overall data for the analysed scene is gathered mainly to determine the genre of the video and matching it with more videos analysed in the database .These matched videos are recommended for the user or it can be used as a search tool for finding related videos in the same genre or having a scene with similar features.

2.1 Objective

The system purpose to increase the accuracy of the videos recommendation to the users on many platforms, So the users will find the more interesting videos to them accurately. also it will save many calculations the platforms uses to recommend a videos to their users.Also a video search feature will be a available for the users to insert a video and get similar videos as a result.

2.2 Scope

The system will work on YouTube Data-set which is the most popular platform also it's supported by an enormous number of videos with different genre. Feature extraction technique will take place to improve the recommendations of the videos to the users.

2.3 Project Overview

In this system, the user will insert a video to our program in certain intervals, where the video will get segmented into a set of frames. The frames will then be sent to the YOLO algorithm, there it will check with the YouTube 8 Million Videos dataset for the similar videos currently existing in the database, then after the video is classified, it will be inserted into the currently existing database as well. The program will then collect data similar to the intervals the user sent and based upon the dataset inserted, it will start handing him more video recommendations or videos resulted from him searching according to the intervals he inserted.

overview.png overview.png

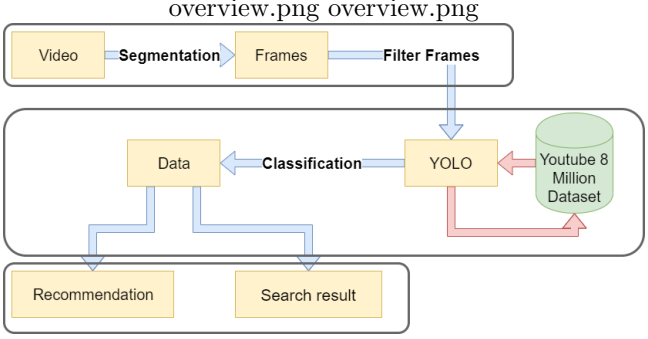

Figure 1: System Overview

3 Similar System Information

1. This paper was motivated to work in the internet multimedia industry, to help with the cold-start problem for new videos, to solve this problem, in this paper we propose a content-based video recommendation approach by taking the advantage of deep convolutional neural networks to alleviate the cold-start problem, and the researchers reached out that with the support from deep convolutional neural networks, frame-level vision feature exhibits its superiority against textual meta-data and audio information, and this helps users discover more contents of interests.

2. The motivation for the work on this paper is to automate the image processing of multimedia, because of the rapird multimedia growth that continues to raise new challenges in the image processing world, and a eliable system is needed for it. The main problem in this work is in video content analysis, as the researchers contributed to the problem by applying systems consisting of three main parts which includes: Shot Boundary Detection, Hierarchical video summarization, retrieve and index target video. the researchers' conclusion then was, Information theory based approach was proposed to video retrieval and summarization problem, as the results show the usefulness of the proposed method.

3. In this paper, the researchers main goal is to improve the recommendation system in YouTube's system, as the recommendations in YouTube are extremely challenging due to the videos' scale, noise and freshness. The researchers solved this by using Google Brain, which was recently open sourced as TensorFlow, with their system using two neural networks. Their approach performed much better on watch-time weighted ranking evaluation metrics compared to predicting click-through rate directly.

4. "Multiple object detection for smart TV shopping video using point to point feature based SURF method.".
   In [2] Motivation: In an attempt to make multiple detection to put every single object for shopping.
   Problem Statement: The major problem to attain this is the multiple detection of complex objects like bags, dresses, accessories and so on, from the content stream of a video. This is more difficult to attain during TV shows, because in general the cameras are not static and the target objects do not have any specific form, as numerous captured angles are possible.
   Contribution: They used Speeded Up Robust Feature(SURF) technique to identify multiple type of objects in a video. It is based on Hessian matrix determinant. The proposed technique is successful in extracting target objects with low error rate.
   Read the data, detect feature points, extract feature descriptors and find putative matched points, locate the object in the scene using putative matches, transformed polygon indicates the location of the object in scene, get the bounding box of the reference image.

5. The researchers were motivated due to the amount of multimedia content being drastically increased over the internet, with the problem of having abundance of semantic information within the data uploaded that's being ignored. The researchers contributed by applying s MAC-REALM, a framework for extraction of syntactic and semantic content features and content modelling with either little or no user interaction, reaching the conclusion that MAC-REALM can convert 'raw media into a content model through a process of content feature extraction and modelling'.

6. The researchers were motivated due to the need of revision for video retrieval technique by content based style, As text based video retrieval is degrading its performance with respect to major issue of user's probabilistic perception to a video. The paper proposes the multilevel Thepade's sorted Ternary Block Truncation Coding (TSTBTC) based video retrieval. Also, the concept of intermediate blocks of video frames is used for the performance improvisation of video retrieval and even-odd videos are proposed to be used for performance rise in CBVR, with the result being: TSTBTC with combination for cases as intermediate blocks, even odd videos and multilevel has shown better performance than binary BTC in all considered color spaces (RGB, KLUV, YIQ, YUV, YCgCb, YCbCr and XYZ).

7. The researchers were motivated due to the increasing amount of video data day by day, in the field of key frame extraction from videos, to fix the problem of browsing and retrieving the videos being increasingly difficult due to the uncertain length and formats of videos, making their

accessing remain a challenge. The researchers proposed novel Thepade's Sorted Ternary Block Truncation Coding (TSTBTC) based key frame extraction techniques with help of seven assorted similarity measures. The paper's result was that detection of key frame among all the frames in a video saves a lot of time and resources in processing of the videos in question.

8. The researchers were motivated due to the urgent need to develop intelligent techniques for object indexing and retrieval in a large-scale video data. The problem statement of the work is the challenging task of developing an efficient object retrieval solution, especially for surveillance videos. In this paper, a mask assisted object encoding approach is presented to boost retrieval performance in surveillance videos. The main results the researchers reached were that with experiments and comparisons, efficiency was demonstrated for the proposed approach.

9. The researchers were motivated due to the the increasing data sizes of datasets available everywhere, making the retrieving process become harder, as it takes more storage, computation speed and shape to represent. In this paper, the researchers propose a deep learning based method for content-based 3D shape retrieval in the large database, using a view-based method, that consists to capture a set of 2D views around the 3D shape of the object, by using a 55-dimension probability prediction vector as the descriptor of a given 3D object. The main results of the researchers reach is that their method has proven very effective compared to other methods.

10. The researchers were motivated as it may benefit customer profile analysis, contextaided people identification, and computer aided fashion design. The problem faced in the paper is that a real-time clothing recognition system, especially for surveillance videos, remains challenging. The researchers contributed to solve the problem by taking advantage of face detection and tracking to locate human figures and develop an efficient clothing segmentation method utilizing Voronoi images to select seeds for region growing. They also compare clothing representations combining color histograms and 3 different texture descriptors. The main result the researchers reached is that the achieves average recall rate 80

3.1 Similar System Description

a system was proposed which aimed to enhance the video experience by users and to maximise the on screen time a user will spend on their platform to maximise profit however this system analysed the video scenes based on three major components, audio , subtitles , and video, which gave a decent idea about what the scene is really about .

3.2 Comparison with Proposed Project

The proposed project will enhance these feature extraction system to get results with greater accuracy , also adding the video search by scene as an exclusive feature using data from the video used for the search and videos analysed form the database

3.3 Screen Shots from previous systems (if needed)

4 Project Management and Deliverables

4.1 Tasks and Time Plan

		Name	Duration	Start	Finish	Predecessors	Resource Names
1		Idea Discussion	11 days?	7/26/19 ...	8/9/19 5:...		
2		Idea Research	23 days?	8/15/19 ...	9/16/19 ...	1	
3		Proposal Evaluation	4 days?	9/25/19 ...	9/30/19 ...		
4		Designing application	3 days?	10/3/19 ...	10/7/19 ...		
5		Implementing Prototype	3 days?	10/8/19 ...	10/10/19...	4	
6		Submittinng Contribution o...	9 days?	11/8/19 ...	11/20/19...		
7		Designing Database	4 days?	10/14/1...	10/17/19...		
8		Class Diagram	2 days?	10/18/1...	10/21/19...		
9		SRS Writing	14 days?	10/22/1...	11/8/19 ...	8	
10		SRS Evaluation	1 day?	11/11/1...	11/11/19...	9	
11		Dataset Collection	6 days?	11/11/1...	11/18/19...		
12		Data Classification	6 days?	11/18/1...	11/25/19...		
13		Implementing Application	20 days?	11/26/1...	12/23/19...	12	
14		External Examiner	1 day?	12/24/1...	12/24/19...	13	
15		SDD Writing	19 days?	12/23/1...	1/16/20 ...		
16		SDD Evaluation	0 days?	1/18/20 ...	1/20/20 ...	15	
17		Validation and Testing	41 days?	1/31/20 ...	3/27/20 ...		
18		Prototype Evaluation	3 days?	2/15/20 ...	2/19/20 ...		
19		Testing System	5 days?	4/16/20 ...	4/22/20 ...	17	
20		Technical Evaluation	5 days?	5/1/20 8...	5/7/20 5:...		
21		Writing Thesis	29 days?	4/21/20 ...	5/29/20 ...		
22		Final Thesis	7 days?	6/20/20 ...	6/30/20 ...	21	

Figure 2: Time Plan

5 References

[1] Covington, P., Adams, J., Sargin, E. (2016). Deep Neural Networks for YouTube Recommendations. Proceedings of the 10th ACM Conference on Recommender Systems - RecSys '16.doi:10.1145/2959100.2959190

[2] Divya Chhabra, Amandeep Verma, "Multiple object detection for smart TV shopping video using point to point feature based SURF method," in 2016 International Conference on Inventive Computation Technologies (ICICT), 2016.

[3] Li, Y., Wang, H., Liu, H., Chen, B. (2017). A study on content-based video recommendation. 2017 IEEE International Conference on Image Processing (ICIP).doi:10.1109/icip.2017.8297150

[4] Yarmohammadi, H., Rahmati, M., Khadivi, S. (2013). Content based video retrieval using information theory. 2013 8th Iranian Conference on Ma-

chine Vision and Image Processing (MVIP).doi:10.1109/iranianmvip.2013.6779981
[5] Parmar, M., Angelides, M. C. (2015). MAC-REALM: A Video Content Feature Extraction and Modelling Framework. The Computer Journal, 58(9), 2135–2171. doi:10.1093/comjnl/bxv042

[6] Thepade, S. D., Subhedarpage, K. S., Mali, A. A. (2013). Performance rise in Content Based Video retrieval using multi-level Thepade's sorted ternary Block Truncation Coding with intermediate block videos and even-odd videos. 2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI). doi:10.1109/icacci.2013.6637306

[7] Thepade, S. D., Patil, P. H. (2015). Novel visual content summarization in videos using keyframe extraction with Thepade's Sorted Ternary Block truncation Coding and Assorted similarity measures. 2015 International Conference on Communication, Information Computing Technology (ICCICT). doi:10.1109/iccict.2015.7045726

[8] Teng, K., Wang, J., Xu, M., Lu, H. (2014). Mask Assisted Object Coding with Deep Learning for Object Retrieval in Surveillance Videos. Proceedings of the ACM International Conference on Multimedia - MM '14. doi:10.1145/2647868.2654981

[9] Benjelloun, M., Dadi, E. W., Daoudi, E. M., Larhmam, M. A. (2018). Content-based 3D shape retrieval using deep learning approach. Proceedings of the International Conference on Learning and Optimization Algorithms: Theory and Applications - LOPAL '18. doi:10.1145/3230905.3230957

[10]Yang, M., Yu, K. (2011). Real-time clothing recognition in surveillance videos. 2011 18th IEEE International Conference on Image Processing. doi:10.1109/icip.2011.6116276

[11] Covington, P., Adams, J., Sargin, E. (2016). Deep Neural Networks for YouTube Recommendations. Proceedings of the 10th ACM Conference on Recommender Systems - RecSys '16. doi:10.1145/2959100.2959190

[12] Yarmohammadi, H., Rahmati, M., Khadivi, S. (2013). Content based video retrieval using information theory. 2013 8th Iranian Conference on Machine Vision and Image Processing (MVIP). doi:10.1109/iranianmvip.2013.6779981

[13] Li, Y., Wang, H., Liu, H., Chen, B. (2017). A study on content-based video recommendation. 2017 IEEE International Conference on Image Processing (ICIP). doi:10.1109/icip.2017.8297150

[14] https://www.sciencedirect.com/topics/engineering/feature-extraction