KLE Society's
KLE Technological University

A Minor Project Report

On

OBJECT BASED VIDEO SUMMARIZATION

submitted in partial fulfillment of the requirement for the degree of

Bachelor of Engineering

In

Computer Science and Engineering

Submitted By

Goutami H	01FE17BCS077
Laxmi Savadi	01FE17BCS101
Malatishree S	01FE17BCS103

Under the guidance of
Dr. SUJATA C

SCHOOL OF COMPUTER SCIENCE & ENGINEERING

HUBLI–580 031 (India).

Academic year 2019-20

KLE Society's
KLE Technological University

2019 - 2020

SCHOOL OF COMPUTER SCIENCE & ENGINEERING

CERTIFICATE

This is to certify that Minor Project entitled **Object based video summarization** is a bonafied work carried out by the student team **Mrs. Goutami H – SRN: 01FE17BCS077, Mrs. Laxmi Savadi - SRN: 01FE17BCS101, Mrs. Malatishree S-SRN: 01FE17BCS103** in partial fulfillment of completion of Sixth semester B. E. in Computer Science and Engineering during the year 2019 – 2020. The project report has been approved as it satisfies the academic requirement with respect to the project work prescribed for the above said program.

Guide **Head, SoCSE**

Dr.Sujata C **Dr. Meena S. M**

External Viva:

Name of the Examiners	**Signature with date**
1.	
2.	

ABSTRACT

Object based Video Summarization aims to summarize the video based on user requirements. The object based video summarization has been proposed to improve faster browsing of large video collections and more efficient content access. The TVSum and SumMe datasets are used to convert the entire whole into summary. The video is given as an input and is processed through the yolo architecture (algorithm) to detect the objects present in the frame. After the detection of objects in the frame, select the frames of the specified object. Summarize the video based on the specified object by selecting the salient frames.As huge amount of multimedia repositories make the browsing, retrieval and delivery of video contents are very slow and difficult tasks. Early video summarization methods were built mainly upon visual qualities ,while the object based video summarization is more abstract and higher-level cues are leveraged in summarization frameworks. The object based video summarization helps the users to view and retrieve the necessary information in the entire video which saves the time of users.

ACKNOWLEDGEMENT

This project is the result of the combined efforts of the entire team. However, it would not have been possible without be kind support and help of many individual and faculty we would like to extend our sincere thanks to all of them.
We are highly indebted to our guide Dr.Sujata C for her guidance and constant super vision has well as providing necessary information regarding the project and also for his moral and technical support in completing the project.as well we would like to thank for the faculty members for the support and guidance throughout the semester. We would also like to thank our Head of SoCSE Dr.Meena S M and Head of SoECE Dr.Uma.
and Vice Chancellor Dr.Ashok Shettar for giving the opportunity to do this project.
well we would like to express our gratitude towards our parents and member of KLE TECHNOLOGICAL UNIVERSITY for their kind cooperation for this project.

Goutami H 01fe17bcs077

Laxmi S 01fe17bcs101

Malatishree S 01fe17bcs103

Chapter 1
INTRODUCTION

1.1 Overview of the project

In recent years, because of the rapid growth in digital video technologies, the information retrieval has become more and more important and lots of research efforts have been devoted to the video retrieval and video analysis based features. Due to the increasing volume of video content on the Web, and the human effort taken to process it, new technologies need to be researched in order to develop techniques to manage effectively and efficiently the huge amount of data.Indeed, both consumers and professionals have the access to ubiquitous video acquisition devices nowadays .While the video data is a great asset for information extraction and knowledge discovery, due to its size and variability ,it is extremely hard for users to monitor or find the occurrences in it.

So to overcome the problem a video summarization algorithms are used which allow us to quickly browse a lengthy video by capturing the essence and removing redundant information.

1.2 Challenges

1. To correctly detect the objects in the video.
2. To detect the salient frames in the video.
3. After detecting the salient frames correctly, we need to summarize the video based on user requirement.
4. Accuracy of the object learning algorithms.
5. Evalute the performance of a video summarizer.

1.3 Motivation

1. The main intention to choose the project is to summarize the video based on the user requirements.
2. Th system designed reduces the viewing of very long duration videos.
3. To increases the efficiency of rereviewing data compared to other video-summarization techniques

1.4 Objectives of the project

- To detect user input objects present in the frame.
- To select the frames of the specified object.
- To select the salient frames for generating the video summarization based on specified object.

.

1.5 Problem definition

The video data is a great asset for information extraction and knowledge discovery. Object-based video summarization system focuses on user input objects and retrieves the summarized video related to the user input object.

1.6 Literature Survey

There are generally two main approaches for Hand Gesture Recognition for HCI (Human Computer Interaction), first is hardware based and second is vision based.

[1]FROM KEYFRAMES TO KEY OBJECTS:Jingjing meng, hongxing wang, junsong yuan1 yap-peng tanschool of electrical and electronic engineering, nanyang technological university, singapore school of software engineering, chongqing university, china,CVPR 2016.
A selection problem is developed in this paper as a sparse dictionary selection problem, i.e. selecting a few present object proposals to recreate the entire proposal collection.A proximal gradient problem and solve it by the fast iterative shrinkage thresholding algorithm.Compare with current set of members
Approaches such as K-mediod, scanty dictionary range and ba sed density
Selection confirm that the primary video artifacts can be better captured by our formulation.

[2] VIDEO CO-SUMMARIZATION: VIDEO SUMMARIZATION BY VISUAL CO-OCCURRENCE
Wen sheng chu1, Yale song ,Alejandro jaimes robotics institute, carnegie mellon university yahoo labs, new york,in 2015.
In this paper deals to develop maximal biclique finding(mbf)algorithm that
to a sparsely co-occurring patterns, discarding less co-occurring patterns even if they are dominant in one video.
Hence,can easily scale upto handle a large number of videos simultaneously

[3] QUERY-FOCUSED VIDEO SUMMARIZATION: Aidean Sharghi, Jacob S. Laurel, and Boqing Gong Center for Research in Computer Vision, University of CentralFlorida, Orlando, FL 32816 ,Department of Computer Science, University of Alabama at Birmingham, AL 35294.
The paper deals with the study of one of the main obstacles to the research on video Summarization:The user subjectivity is description — users have different biases regarding summaries. The subjectivity causes two problems, at least. First, no single video summarizer suits all users unless it communicates with the individual users and adapts to them.
Second, it is very challenging to evaluate the performance of a video summarizer.
To tackle the first question, we explore the recently proposed query-focused video summary which introduces user preferences into the summary process in the form of text queries about the video. To answer the second issue, we argue that a successful assessment criterion for video summarization would concentrate on the semantic details that can be interpreted by humans rather than the visual or temporal overlaps. To this end, we collect dense concept annotations per video-shot, compile a new dataset, and propose an efficient evaluation method based on the concept annotations.

Chapter 2

PROPOSED SYSTEM

2.1 Description of proposed system .

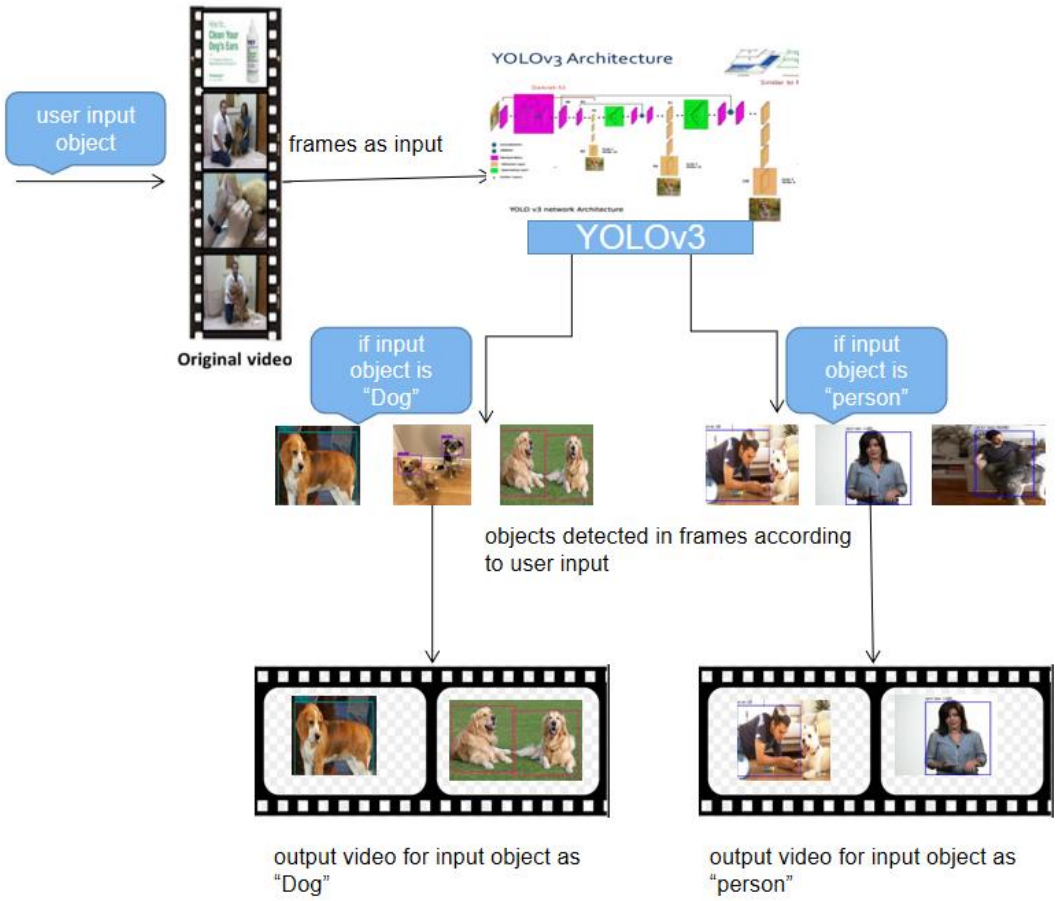

Figure 1:Description of proposed system

The project aims to summarize the video based on user requirements. The given input video is converted into set of frames and these frames are processed through the Yolo architecture (algorithm) to detect the objects present in the frame. to detect the object in the frame parameter such as probability was used, for example, if the probability of finding an object in the frame is greater than 30% than the object is detected else the object is not detected in the frame this also helps us in extracting the silent frames present in the video. After the detection of objects in the frame, select the frames of the specified object. By storing frames into an array, later on a set of frames in an array are converted into a video.

2.2 Description of Target Users

The project can be used for office, school, shops etc. for easy usage purposes. The video data is a great asset for extracting information and discovery of knowledge. Due to its size and variability, it is extremely hard for users to monitor or find occurrences in it. One of the main obstacles of video summarization is the user subjectivity. This project overcomes the problem by summarizing the video based on user queries.

2.3 Advantages/applications of proposed system

- Video summarization helps the user to quickly browse a lengthy video by capturing the essence and removing redundant information.
- Video summarization helps in efficient storage.
- To provide the user a synthetic and useful visual abstract of video sequence, a short summary of video is required.

2.4 Scope

Video summarization opens a wide branch in compact representation of the multimedia data. Based on the application different type of summary could be derived from the videos. It could be revolutionary in the field of communication for frame dropping technique. In the future the implementation of the algorithm to the real could be possible. For the motion detection, motion vector estimation and motion compensation could be used which is more advanced and could speculate for the better subjective readings. Introduction of QoE in video summarization is new concept and provides better evaluation of the algorithm.

Chapter 4

SYSTEM DESIGN

4.1 Architecture of the system

- An architecture diagram is a graphical representation of a set of concepts , that are part of an architecture , including the their principles , elements and components.

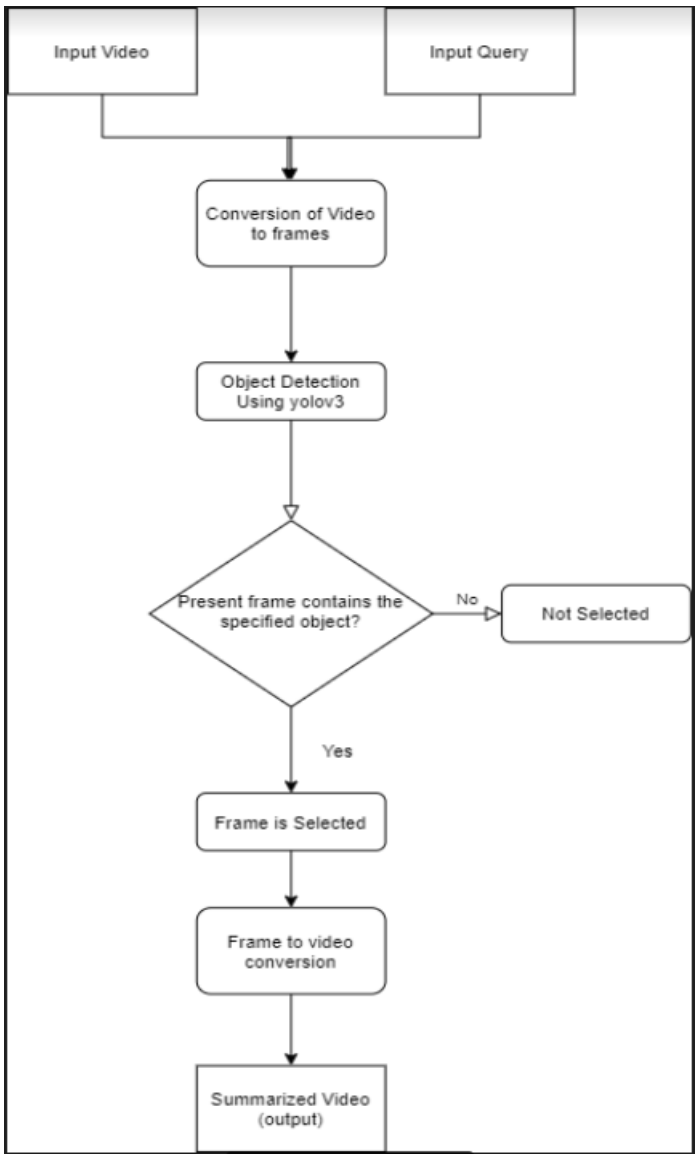

Figure 2:Architecture Diagram

Chapter 5

IMPLEMENTATION

5.1 Proposed methodology

The project aims to summarize the video based on user requirements. The video is given as a input and is processed through the yolo architecture(algorithm) to detect the objects present in the frame. After the detection of objects in the frame, select the frames of the specified object. By storing frames into an array, later on a set of frames in an array are converted into a video.

5.2 Description of Modules

Yolo Architecture

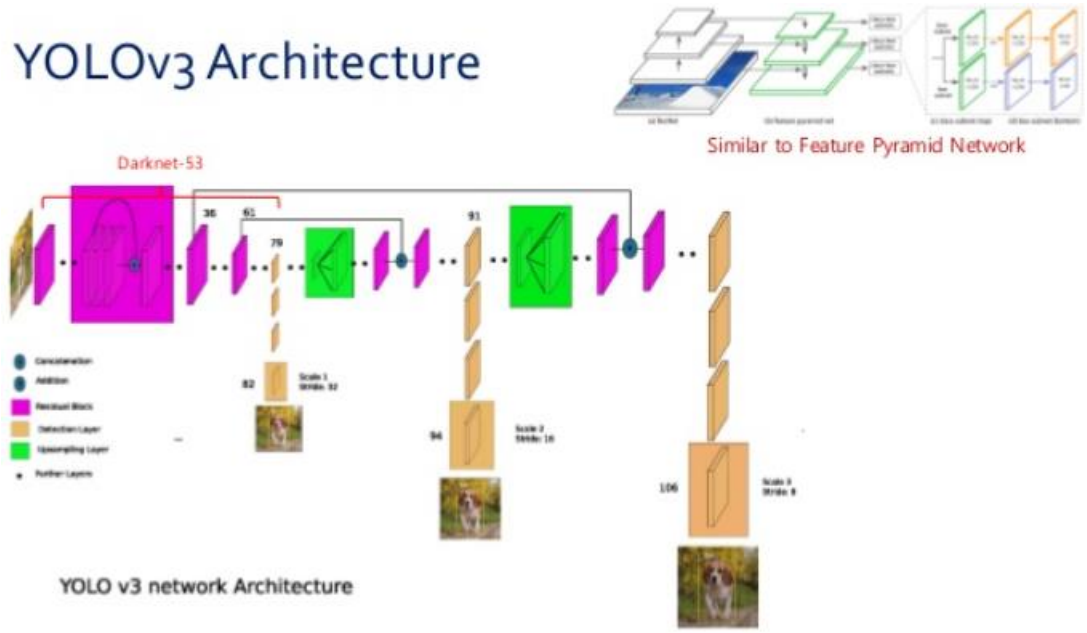

Figure 3:Yolov3 Architecture

YOLO is a convolutionary neural network built to do real-time object detection. The algorithm applies the entire image to a single neural network, then divides the image into regions and predicts bounding boxes. The most prominent characteristic of v3 is that it makes three different scales of detections.YOLO is a completely convolutionary ne twork and it produces its eventual output by adding a 1 x 1 kernel to a function map. For YOLO v3, the detection is achieved by applying 1 x 1 detection kernels to three-size function maps at three different locations in the network.The video was mainly conve rted to frames than processed for object detection in frames through the YOLO v3. The object within the frames is identified by the user input query. Than object identified f rames (salient frames) are counted and translated into video which is also called the conde nsed video.

Detection of objects.

Figure 4 : bicycle detected

Figure 5:Person detected

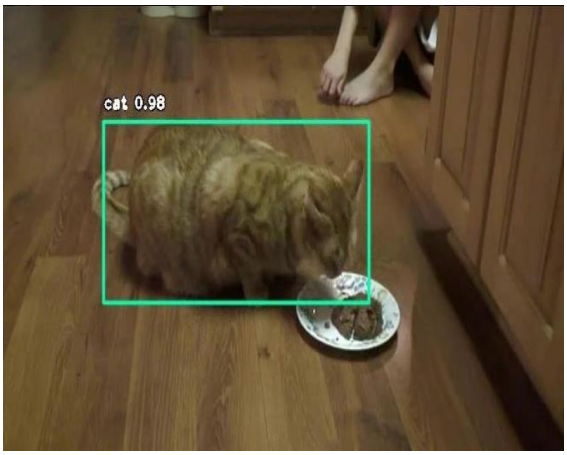

Figure 6:cat detected

Chapter 6

TESTING

Method	Dataset	F1-Score
Video Summarization with Attention-Based Encoder-Decoder Networks	SumMe	44.4
	TVSum	61.0
Unsupervised Video Summarization with Adversarial LSTM Networks	SumMe	41.7
	TVSum	56.3
Query-Focused Video Summarization: Dataset, Evaluation, and a Memory Network Based Approach	UT Egocentric	62.66
Object-Based Video Summarization (ours)	SumMe	69.7
	TVSum	75.6

Chapter 7

CONCLUSION

The video summarization is intended to afford comprehensible analysis of video.The framework has been actualized in PYTHON environment. The TVSum and SumMe datasets are used to convert the entire whole into summary based on user choice. It helps the users to view and retrieve the necessary information in the entire video which saves the time of users. The innovation has wide applications in the fields of expanded reality, business field, schools, offices etc. The intention was to make innovation in the least expensive conceivable way and furthermore to make it under an institutionalized working framework.

Chapter 8

REFERENCES

[1] Query-Focused Video Summarization: Dataset, Evaluation, and A Memory Network Based Approach Aidean Sharghi†, Jacob S. Laurel‡, and Boqing Gong††Center for Research in Computer Vision, University of Central Florida, Orlando, FL 32816

[2]Video Co-summarization: Video Summarization by Visual Co-occurrence Wen-Sheng Chu1 Yale Song2 Alejandro Jaimes2 1Robotics Institute, Carnegie Mellon University 2Yahoo Labs, New York,CVPR 2015

[3]YOLO:http://www.youtube.com/playlist?list=PLXLrBk6h3wSGvuTnxB2Kj358XfctL4BM

[4] From Keyframes to Key Objects: Video Summarization by Representative Object Proposal Selection
Jingjing Meng1 Hongxing Wang1,2 Junsong Yuan1 Yap-Peng Tan1 1School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 2School of Software Engineering, Chongqing University, China,CVPR 2015