Video Summarization using Genetic Algorithm

A Report Submitted
in Partial Fulfillment of the Requirements
for the Degree of
Bachelor of Technology
in
Computer Science & Engineering

by

Anubhav Rajput (20194037)
Anurudh Pratap Singh (20194067)
Divyanshi Agrawal (20194171)
Gaurav Dalal (20198069)

under the guidence of
Dr. Ranvijay

to the

COMPUTER SCIENCE AND ENGINEERING DEPARTMENT
MOTILAL NEHRU NATIONAL INSTITUTE OF TECHNOLOGY
ALLAHABAD, PRAYAGRAJ
May, 2023

UNDERTAKING

I declare that the work presented in this report titled "*Video Summarization using Genetic Algorithm* ", submitted to the Computer Science and Engineering Department, Motilal Nehru National Institute of Technology Allahabad, Prayagraj, for the award of the ***Bachelor of Technology*** degree in ***Computer Science & Engineering***, is my original work. I have not plagiarized or submitted the same work for the award of any other degree. In case this undertaking is found incorrect, I accept that my degree may be unconditionally withdrawn.

May, 2023
Allahabad

(Anubhav Rajput 20194037)

(Anurudh Pratap Singh 20194067)

(Divyanshi Agrawal 20194171)

(Gaurav Dalal 20198069)

CERTIFICATE

Certified that the work contained in the report titled "*Video Summarization using Genetic Algorithm*", by *Anubhav Rajput, Anurudh Pratap Singh, Divyanshi Agrawal, Gaurav Dalal*, has been carried out under my supervision and that this work has not been submitted elsewhere for a degree.

(Dr. Ranvijay)
Computer Science and Engineering Dept.
M.N.N.I.T, Allahabad

May, 2023

Preface

Nowadays, there is a huge amount of information available online in the form of lengthy videos making it difficult for consumers to rapidly find the information they need. Automated video summarising approaches are necessary to efficiently search and gather useful information from these lengthy films.

The summarization of the segmented video is a critical stage in the process of producing video thumbnails, video surveillance, and video downloads. When summarising a video, a few frames from each scene are combined to create a short film that briefly summarises the entire video's plot. The proposed research project covers the segmentation and synthesis of the frames. A genetic algorithm (GA) for segmentation and summarization is used to observe the highlight of an event by selecting a few significant frames.

This report contains different sections like the needs of this project, technologies used, the analysis of the results, algorithms applied, and its overall impact. We have also made the analysis of the results of video summary obtained by genetic algorithm using F-Score.

Acknowledgements

The completion of this project required a lot of effort, guidance, and support from many people. We feel privileged and honored to have got this all along with the development of the project. We would like to express our gratitude to our mentor, Dr. Ranvijay for his perennial support, guidance and monitoring throughout the making of this project. We would also like to acknowledge and thank our professors, colleagues, and seniors for supporting us and enabling us to successfully complete our project.

We would also like to express our sincere gratitude to Prof. R.S. Verma, Director, MNNIT Allahabad, Prayagraj, for providing us with the facilities to complete the project.

Contents

Chapter 1

Introduction

Video lectures, personal videos, documentaries, sports films, and movies in other fields are becoming the most common and practical way to transmit information. It takes a lot of time, effort, and hardware storage to process these enormous volumes of footage. Video summaries are essential in this circumstance.

Even for us humans, summarising videos is a challenging task by nature. One needs to watch the complete video material in order to choose the most significant sections, subject to the desired summary length. That's why video summarization is really necessary. But video summarization depends on user perspective.

The aim of video summarization is to produce a fluent summary of the provided original video, highlighting the most important details while removing repetition from other sources. In this thesis, we propose video summarization using genetic algorithm which takes motivation from "Video segmentaion and summarization using modified genetic algorithm"[8]

1.1 Types of video summarization

1. Keyframe-based video summarization: Using this method, a summary of the video is made by choosing a few representative frames that represent the vital information.

2. Video skimming: Video skimming is a method that involves selecting a subset

of the video frames that most accurately depict the key scenes.

3. Object-based video summarization: It identifies certain video objects, such as people or cars, and producing a summary based on their interactions.

4. Motion-based video summarization: Motion-based video summary analyzes the film's motion patterns and chooses the frames based on motion patterns.

5. Semantic-based video summarization: Using computer vision and natural language processing techniques, this technique analyses the video's meaning and creates a summary based on its material.

In this paper, we propose genetic algorithm based video summarization system to extract key frames to generate a video summary of original video. The proposed approach is key frame based video summarization system that also focuses on identifying motion patterns and specific objects.

1.2 Motivation

Video summarization enables user to have better browsing and searching experience. It also reduces the storage requirements for video data thus reducing cost and energy consumption.

Genetic algorithm is helpful as it can automate the process of selection important frames from the long videos based on some specific criteria.

The following are the benefits of using a genetic algorithm.

1. Genetic algorithms are capable of handling large datasets, and they can find ideal answers quickly.

2. Genetic algorithms are an automated optimization technique, which can find optimal solutions without human aid.

3. Genetic algorithms do not require prior knowledge about the video.

4. Genetic algorithms are flexible and can adapt to different types of video content.

Chapter 2

Preliminaries

2.1 Randomised Algorithm

A randomised algorithm produces probabilistic results by using randomization in some of its parts. This means that a randomised algorithm may provide different results each time it is performed, even with the same input.

Applications for randomised algorithms include producing random numbers, database searches, solving optimisation problems, and determining whether a program is accurate.

2.2 What are Nature Inspired algorithms?

The nature-inspired algorithms work in the same way as natural processes including evolution, swarm behaviour, and neural networks. The main idea is to use the principles of natural systems to solve difficult optimisation challenges.

Artificial neural networks, genetic algorithms, particle swarm optimization, ant colony optimization, and differential evolution are a few examples of algorithms that draw inspiration from nature.

2.2.1 Genetic Algorithm

An example of an optimisation algorithm that follows natural selection procedures are genetic algorithms. In genetic algorithms, the fittest people have a higher probability of surviving and transferring their genetic information to the following generation. Less fit people, on the other hand, have a lesser chance of surviving.

The process of genetic algorithms can be broken down into several steps, as follows:

1. Initialization: A group of potential solutions is generated randomly with size defined by the user.

2. Evaluation: The fitness of each solution is calculated by the fitness function.

3. Selection: The most physically fit members of the population are chosen to have children. Some selection methods are :-

   (a) Roulette Wheel Selection: This is a common method of selection where each member of the population is given a probability of selection based on how fit they are. A higher level of fitness increases the likelihood that a person will be chosen for the optimisation process.

   (b) Tournament Selection: This method compares the fitness levels of a small sample of people it randomly selects from the population. The parent chooses the one with the greatest fitness rating.

4. Crossover: Through a crossover operation, the chosen parents are combined to produce new offspring.

   (a) Single Point Crossover: This is the most basic sort of crossover, in which a single spot on the chromosomes of the two parents is picked at random, and the genetic material is exchanged beyond that point to produce two new children.

   (b) Two Point Crossover: Using this technique, two crossover points are selected, and the genetic material between the two locations is exchanged to produce two new children.

(c) Uniform Crossover: This method involves randomly selecting each bit from one of the parents to create the offspring.

5. Mutation: Some of the offspring may be randomly mutated to introduce new genetic material into the population.

   (a) Bit Flip Mutation: A random bit in the chromosome is flipped from 0 to 1 or vice versa using this fundamental mutation operator.

   (b) Swap Mutation: In this procedure, two genes are chosen at random and their locations are switched.

   (c) Inversion Mutation: In this procedure, a chromosome section is chosen, and the order of the genes on that segment is reversed.

6. Replacement: The old population is replaced with the new population of offspring.

7. Termination: The algorithm stops after a fixed number of iterations.

2.3 YOLO

YOLO (You Only Look Once) is an object detection algorithm that was first introduced in 2016 by Joseph Redmon, et al [4]. YOLO is a deep learning system. It identify objects present in images or videos by dividing the input into grid cells. It then predicts the bounding box and class probabilities for each of these cells. The bounding box identifies the object's rectangular location within the image, and the class probabilities indicate which classes the object is most likely to fall under.

The basic steps involved in the YOLO algorithm are as follows:

1. Image Input: The first step is to input an image or video frame into the algorithm.

2. Grid Division: YOLO predicts the bounding box and class probability for each cell by dividing the image into grid of cells.

3. Bounding Box Prediction: It predicts the cordinates of bounding box containing an object in each cell.

4. Object Classification: It also predicts ts the confidence score of object belonging to particular class.

2.4 OpenCV

The open-source library known as OpenCV, or Open Source Computer Vision, has a huge application in computer vision and machine learning techniques. Numerous industries, including robotics, automation, security, and image and video processing, heavily rely on the library.[7]

A wide variety of tools and features are available with OpenCV. The collection includes algorithms for object detection, object tracking, stereo vision, image segmentation, feature matching, and other tasks.

2.4.1 Human detection model using OpenCV

OpenCV provides a wide range of tools and functions for object detection, including human detection. A pre-trained person detection model from OpenCV is based on the SVM classifier and HOG (Histogram of Oriented Gradients) descriptor. This model, which is part of OpenCV's "cv2" module, can be used to find people in pictures or videos.

Five HOG filters i.e. front looking, left looking, right looking, front looking but rotated left, and a front looking but rotated right are used to construct the model. The pre-trained model must first be loaded into memory using the "cv2.HOGDescriptor()" method before it can be used.

The bounding boxes of the recognized humans, along with their confidence scores, are returned by this model after receiving an image or video frame as input.

Chapter 3

Related Work

Recent years have seen major advancements in the study of video summarization, particularly with the appearance of deep learning-based methods. Due to their capacity to automatically extract important elements from video data and produce precise summaries, these techniques have grown in popularity.

There are two main approaches to video summarization: supervised and unsupervised.Significant research has been done on video summarization using unsupervised learning by Kaiyang Zhou, Yu Qiao, and Tao Xiang. Deep Reinforcement Learning for Unsupervised Video Summarization with Diversity-Representativeness Reward was the topic of a paper they presented in 2017. [5].

Using supervised data, the tool generates summaries for other videos belonging to the same category/class. Basak, Jayanta and Luthra, Varun and Chaudhury, Santanu [2] derived features related to frame transitions and then represent each transition as a state. They formulated a loss functional to measure the difference between state transition probabilities in the original video and those in the summary video, and then optimized this function.

This paper describes a method for creating a summary of a video using Genetic Algorithm. The algorithm uses specific techniques to search through the video and identify the most important parts to include in the summary.A journal is published by Yang, Xue and Wei, Zhicheng in 2012 on video summarization using gentic algorithm (Journal of Convergence Information Technology)[13].

Algrorithm we used here involves the use of Real-Time Object Detection Using YOLO to detect objects in keyframes for title based video summarization [4].

Manasa Srinivas, M. M. Manohara Paiand Radhika M. Pai [12] utilized a Rank Based Approach to extract keyframes from a video using features like quality, representativeness, uniformity, static and dynamic attention. These keyframes can be used both for summarizing and indexing purposes. By assigning weights to the features based on their standard deviation, the authors were able to give higher importance to the features with more variation across the frames. This approach helps in identifying the most relevant keyframes.

Keyframe extraction is a prerequisite needed for content-based video indexing, browsing, and retrieval. The primary purpose of keyframe extraction is to manage large amount of video data by selecting a distinct set of frames that are representative of the essential activities in the original video. This simplifies video analysis and processing. The development of keyframe extraction techniques has impacted various fields such as e-learning, news broadcasting, home videos, sports, movies, and many others. [9].

Video skimming, also known as dynamic video summarization, is a method that generates a shortened version of a video by identifying significant components in uni-modal or multi-modal features extracted from the video. A method of soft computing-based text summarising that benefits from linguistic and statistical approaches was presented by Smith, Michael and Kanade, Takeo in 1998 [10]. Its objective is to create a summary that captures the most important moments or scenes in the video while maintaining its temporal connectivity.

In recent years, deep learning techniques have led to significant advances in video summarization. A comprehensive survey by Apostolidis, Evlampios and Adamantidou [1] aims to provide an overview of the latest deep-learning-based methods for generic video summarization.Li et al. (2006) [6] conducted an early study on video summarization approaches and categorized them into two main types: utility-based methods and structure-based methods. The former rely on attention models to identify the most salient objects and scenes in a video, while the latter utilize the inherent structure of video shots and scenes to generate a summary.

Chapter 4

Proposed Work

Video Summarization system using genetic algorithm has various sub tasks like frames extraction from video, preprocessing, object detection, fitness evaluation ,selection, crossover, mutation and finally choosing the key frames on the basis of fitness function after iterating over the number of generations. We have proposed our methodology for three components of video summarization.

1. The frames having objects with higher similarity to the title have higher probability of being included in the final summary.

2. The less is the similarity within content of summary, the more is the diversity of information.

3. Ratio of length of original video and length of summary, also known as Compression ratio.

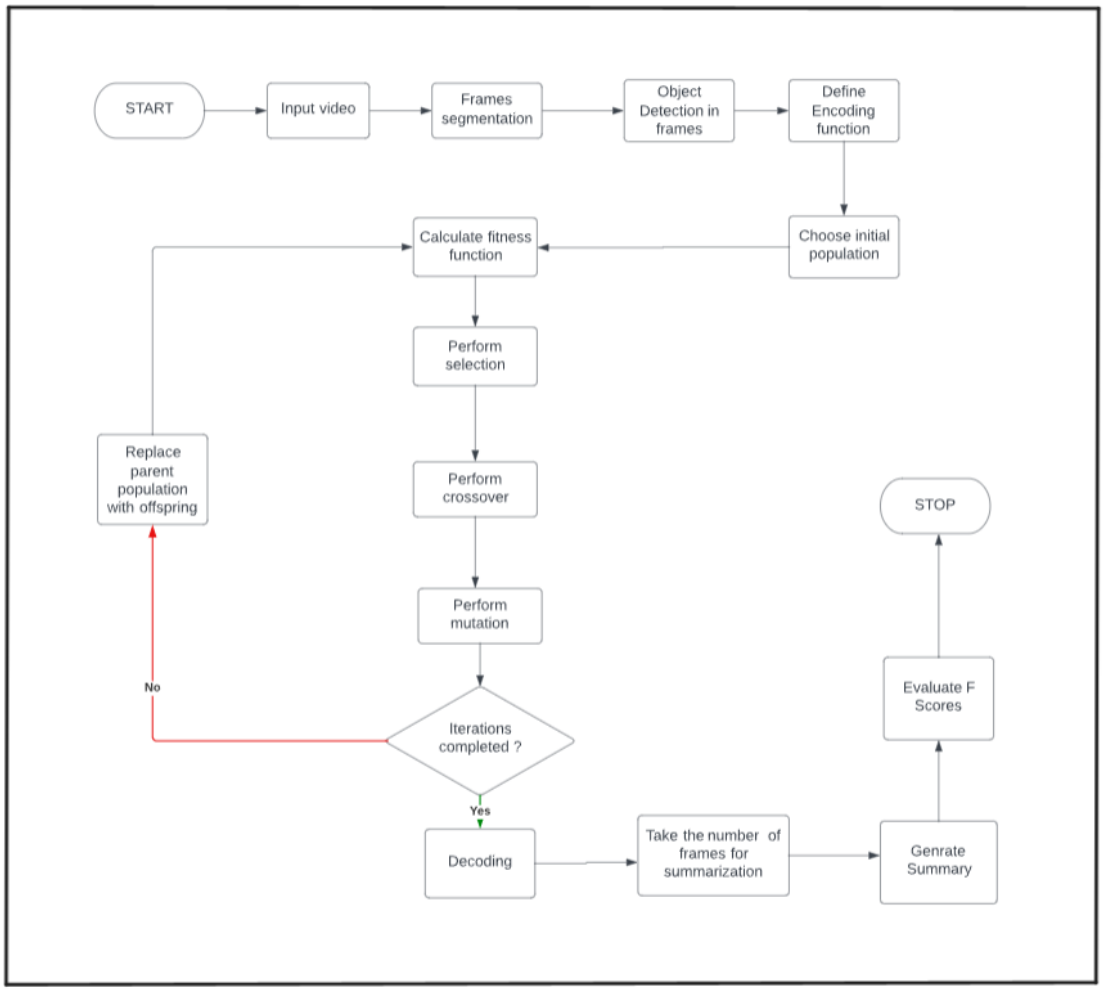

Figure 1: Work flow

4.1 Input pre-processing

Some pre-processing steps that are required to be performed for video summarization are :

1. Video segmentation: The first step in video summarization is to divide the video into smaller segments, such as shots or frames. We have used Open-CV for this,which is a popular library for computer vision and image processing in Python.

2. Object detection: Our model focuses upon finding frames that are more relatable to the object. So object detection in frames is really important. YOLO(You Only Look Once) model is used for object detection.

   A single neural network is used by the object detection method YOLO to identify objects in an image or video. The YOLO algorithm gives each bounding box a class label and a confidence score.

   YOLO also classifies objects them into different categories such as cars, people, or animals.

   We have classified objects into 80 different categories.

3. Frame processing: Once objects have been extracted from the video, the next step is to extract relevant features and color pixels from them.

   - All frames are adjusted to particular width and height.
   - All the frames are converted to gray scale as working on gray scale images is faster.
   - The visual features like colors are extracted.

4.2 Summary generation process using Genetic Algorithm

Genetic algorithm includes defining a fitness function, creating initial population, performing selection, crossover, mutation over a number of generations and then choosing best solution on the basis of fitness function.

4.2.1 Analogy

Various analogies are used in this paper for genetic algorithm in context of video summarization :

- Chromosome: It is a set of genes representing a summary of video. It is a binary vector having 0 and 1, where 1 and 0 representing a frame is present or not respectively.
- Gene: Each bit of chromosome vector is a gene which represent a presence state of keyframe in the video. It is either 0 or 1.
- Population: It represent subset of solutions where each solution is a summary of provided video. Out of these solutions, best one is chosen after some number of iterations of genetic algorithm.
- Generation: Generation represent an one iteration of selection,crossover and mutation of video summary.
- Fitness Function: Fitness function is used to evaluate the quality of a summary, based on factors such as how well it represents the key content of the video or how concise it is.

4.2.2 Steps involved

The following steps are taken for video summarization using genetic algorithm:

1. Preprocessing: The first step is to preprocess the video data to extract relevant features that can be used by the genetic algorithm. This involves segmenting video into frames,extracting the features like color pixels and performing object detection using YOLO model.
2. Chromosome encoding: The chromosomes are created by randomly selecting video frames. It is basically a binary vector having length equal to the number of frames in original video and gives information about the presence state of a frame in summary.

   For example, if the chromosome is 1001001001.

   It represents a video of 10 frames in which 1,4,7 and 10 frame is included in the summary.

3. Choosing initial population: The population size considered in our proposed model is 30. The size of final summary is 15% of the original summary.

   So, the initial population has 30 binary vectors of length equal to number of frames in original video.

   Each frame having n number of 1's placed at random position, where n is the length of final summary in terms of frames.

4. Fitness evaluation: The fitness function is used to evaluate the quality of each chromosome in the population. For video summarization, the fitness function is based on few parameters like adjacent dissimilarity score analyzing the motion patterns in the video, human score , similarity of frames with the title etc.

   The similarity of frames with title is found by finding the objects in the frame and then finding the cosine similarity of object vector with the title vector.

5. Selection: The selection step involves choosing the best summary from the population to form the next generation. This is done using binary tournament selection method. The goal is to select the summary with the highest fitness values.

   The steps involved in binary tournament selection method are :

Algorithm 1 Binary Tournament Selection

```
for (i=0 to i < population_size) do
    select two summaries randomly
    calculate fitness score of the two summaries
    choose the the higher fitness value summary for next generation
end for
```

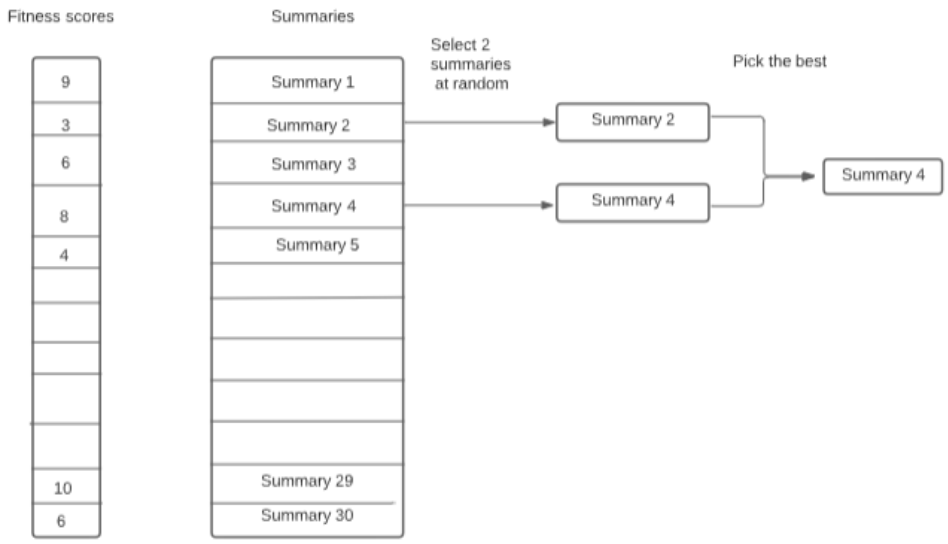

Figure 2: Binary tournament selection

6. Crossover: Crossover is the process of combining two summaries to create new offspring summaries. For video summarization, crossover involves swapping segments between the two parents.

   Single point crossover is used in this proposed work.

   In single point crossover, a random number is chosen in range (0 to n) where n is the length of summary. Then the corresponding segments of 2 parent summaries are swapped with respect to crossover point to form new summaries

Algorithm 2 Single Point Crossover

1: **for** (i=0 to i < population_size) **do**
2: Select two summaries(parents) randomly
3: Choose a random number between 0 and n ▷ n is number of frames in summary
4: Bisect the frame sequence of summary in two segment at chosen number
5: Generate offspring summary by swapping the segments of parent summary
6: **end for**

Figure 3: Single point crossover

In above example , the random crossover point chosen in 2, and segments (0-2) and (3-6) are swapped.

7. Mutation: Mutation involves adding or removing frames or segments from a summary. Mutation adds diversity and exploration to video.

   Bit flip mutation is used in the proposed work.

Algorithm 3 Bit Flip Mutation

1: Select a summary(parent) to mutate
2: Initiate mutation_factor P_m ▷ we considered 0.5
3: **for** (each frame in summary) **do**
4: Choose a random number between 0 and 1
5: **if** $randomNumber \geq mutation\_factor$ **then**
6: flip corresponding bit of frame ▷ invert inclusion of frame in summary
7: **end if**
8: **end for**

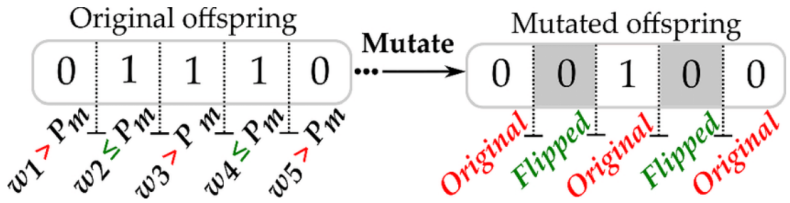

Figure 4: Bit flip mutation

[3]

Here P_m is mutation factor (0.5 considered). For every frame,a random number w_i between 0 and 1 is chosen,and if w_i is greater than P_m, bit is flipped.

8. Repeat: Steps 3-6 are repeated for a fixed number of generations (10). The old population is replaced with the new population of off- spring.

9. Output: After 10 generation, the summary having highest fitness score is the desired summary. The 1's in binary vector corresponds to the frames that need to be chosen and final summary is made of that binary vector.

4.2.3 Fitness Function

Fitness function for genetic algorithm used in this report mainly depends on three parameters human weights, similarity of objects in frames with provided title and difference between adjacent keyframes(genes).

- Human score : This paper primarilty focuses on CCTV surveilance videos and summaries based on activity detection for very long videos, so keyframes with the presence of the humans will have more weight. For detecting human in keyframes, HOG (Histogram of Oriented Gradients) algorithm is used. The algorithm uses the gradients of the image to represent the local appearance and shape of the object being detected. If human is present in the keyframe then human weight 1 is assigned else 0 assigned. Total human_score for a chromosome(summary) will be sum of human weight for all n frames of a summary.

$$human\_score = \sum_{j=1}^{n} human\_weight(F_j) \tag{1}$$

Where

· n is the total number of frames

· F_j is frame under consideration

- Similarity score: For summary that effectively conveys the important information from the source video while also aligning with the given title, we detected objects present in each keyframe using YOLO object detecting model. Then cosine similarity of the object vector of j^{th} frame of chromosome is calculated with the title vector using spacy English language model named "en_core_web_lg" for each keyframe. Similarity score of chromosome will be sum of similarity of all the keyframe(gene) with the title.

$$similarity\_score = \sum_{j=1}^{n} cosine\_similarity(objects(F_j), title) \quad (2)$$

Where

· n is the total number of frames

· F_j is frame under consideration

· title is topic of the video

- Adjacent dissimilarity score : To include portion of video in the summary which have more activity ,difference between adjacent keyframes needs to be calculated. For this difference between the color histogram of pixels of every consecutive frame is calculated and sum of that value for all the keyframes will be the adjacent dissimilarity score.

$$adjacent\_dissimilarity\_score = \sum_{j=2}^{n} dissim(F_j, F_{j-1}) \quad (3)$$

where

· n is the total number of frames

· F_j is frame under consideration

· dissim(F_j,F_{j-1})=pixels[F_j] − $pixels[F_{j-1}]$

Finally the fitness of the solution is calculated as

$$F = (\alpha * HS) + (\beta * SS) + (\gamma * ADS) \quad (4)$$

where

- α, β, γ are weighing factors.
- F is Fitness score
- HS is human score
- SS is similarity score
- ADS is adjacent dissimilarity score

4.3 Algorithm Proposed for video summarization

Algorithm 4 Genetic Algorithm

1: Initialize population
2: **while** (current iteration(itr) < Maximum iteration(Max_itr) **do**
3: Find fitness of population.
4: Select parent summary on basis of binary tournament selection.
5: Perform single point crossover.
6: Perform bit flip mutation on basis of mutation factor.
7: Replace parent population with offspring.
8: $itr \leftarrow itr + 1$
9: **end while**
10: **return** `best summary in the final population.`

Chapter 5

Experimental Setup and Results Analysis

This section illustrates the experimental setup for proposed video summarization system and performance analysis. All the experiments and implementation are done in windows system.

5.1 Experimental setup

For experimental analysis, we have used the following hardware and software specifications:

The software specifications of the setup are as follows:

- python=3.11.2
- VS Code
- jupyter notebook

The hardware specifications of the setup are as follows:

- Processor : Intel(R) Core(TM) i5
- RAM : 8 GB DDR4 Memory

- Storage : SSD

The following libraries of the python has been used :

- numpy
- matplotlib
- spacy
- OpenCV
- YOLO
- scikit-learn

5.2 Dataset

The proposed framework is evaluated on Title Based Video Summarization (TVSum) dataset which is benchmark content for video summarization. It consists of 50 videos from the TVSum along with 50 thumbnail images and titles corresponding to each video. Each video has manual annotations by 20 judges which produces 20 reference summaries for each video. The videos are divided into 10 categories each having 5 videos. The categories are:-[11]

- changing Vehicle Tire (VT)
- getting Vehicle Unstuck (VU)
- Grooming an Animal (GA)
- Making Sandwich (MS)
- ParKour (PK)
- PaRade (PR)
- Flash Mob gathering (FM)

- BeeKeeping (BK)
- attempting Bike Tricks (BT)
- Dog Show (DS).

5.2.1 Reference summaries

The dataset provides us with reference summaries in form of scores (1 to 5) for every 2 second shot of the video. We have 20 responses corresponding to 20 judges for each of the 50 videos. Based on these scores, the best frames are selected and reference summary can be represented as a binary vector with size equal to the number of frames in original video and having 1 or 0 depicting whether frame is present or not respectively.

5.3 Controlling parameters

In any optimisation problem,controlling parameters are based on the application. Controlling parameters used in Genetic Algorithm are used as follows.

Population size	30
Generations	10
Summary size	15%
Mutation factor	0.5

5.4 Result analysis

F-score is a popular metric to assess how well video summarising algorithms work. F-score measures the balance between precision and recall, which are calculated

based on the number of true positives, false positives, and false negatives.

For each of the 50 videos ,average F-score has been calculated of binary vector corresponding to every model generated summary with 20 reference summaries binary vector obtained from the dataset.

As F-Scores are calculated upon the binary vectors of the reference summary and the model generated summary. Thus, for two completely different videos ,the result needs to be significantly low.The observed F-scores for 2 different videos came to be about 0.04.

Recall

Recall measures how well the algorithm included all the crucial and important elements of the original video in the summary.

Recall calculates the ratio of true positives (TP) to the sum of true positives and false negatives (FN). False negatives are frames that are in the reference summary that are not selected by the algorithm, whereas true positives are frames that the algorithm correctly selects.

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative}$$

Precision

Precision is a metric that measures the accuracy of the summary generated by an algorithm.

Precision calculates the ratio of the number of true positives (TP) to the sum of the number of true positives and false positives (FP). False positives are frames or segments that are successfully selected by the algorithm but are not in the reference summary. True positives are frames or segments that the algorithm correctly selects.

The formula for precision is:

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive}$$

F score

F-score measures overall performance of an algorithm. It is the harmonic mean of precision and recall providing an overall measure of the model's performance.

$$FScore = \frac{2 * Precision * Recall}{Precision + Recall}$$

A high F-score indicates that the algorithm has achieved a good balance between precision and recall, and has selected frames or segments that are both relevant and representative of the original video.

5.4.1 Results

The F1-scores of 50 video summaries belonging to 10 categories(5 videos each) was calculated with respect to reference summaries .

Categories	Video 1	Video 2	Video 3	Video 4	Video 5	Average
Category 1 (VT)	0.326	0.320	0.341	0.340	0.341	0.334
Category 2 (VU)	0.316	0.336	0.322	0.337	0.351	0.332
Category 3 (GA)	0.329	0.326	0.348	0.334	0.329	0.333
Category 4 (MS)	0.336	0.324	0.340	0.332	0.320	0.330
Category 5 (PK)	0.320	0.349	0.334	0.321	0.337	0.332
Category 6 (PR)	0.347	0.335	0.332	0.313	0.333	0.332
Category 7 (FM)	0.339	0.326	0.328	0.363	0.329	0.337
Category 8 (BK)	0.330	0.344	0.334	0.316	0.324	0.330
Category 9 (BT)	0.324	0.323	0.328	0.313	0.340	0.326
Category 10 (DS)	0.325	0.340	0.335	0.330	0.329	0.332

Figure 5: F1-scores of 50 videos

We also compared our results with various benchmark models for video summarization. The models include Uniform Sampling(SU), Random Sampling(SR), k-means clustering (CK), spectral clustering (CS), LiveLight(LL), Web Image Prior(WP), Archetypal Analysis from video data only(AA1) and Archetypal Analysis from video and image data(AA2).[11]

	SU	SR	CK	CS	LL	WP	AA1	AA2	GA
Category 1 (VT)	0.39	0.29	0.33	0.39	0.47	0.36	0.33	0.38	0.334
Category 2 (VU)	0.43	0.31	0.40	0.37	0.52	0.48	0.36	0.36	0.332
Category 3 (GA)	0.32	0.36	0.37	0.39	0.46	0.35	0.28	0.33	0.333
Category 4 (MS)	0.37	0.32	0.34	0.39	0.45	0.40	0.36	0.32	0.330
Category 5 (PK)	0.36	0.32	0.34	0.39	0.49	0.27	0.35	0.39	0.332
Category 6 (PR)	0.38	0.30	0.34	0.38	0.42	0.36	0.37	0.42	0.332
Category 7 (FM)	0.32	0.30	0.33	0.37	0.42	0.39	0.41	0.38	0.337
Category 8 (BK)	0.34	0.32	0.34	0.38	0.44	0.41	0.28	0.28	0.330
Category 9 (BT)	0.36	0.29	0.35	0.46	0.45	0.32	0.24	0.27	0.326
Category 10 (DS)	0.33	0.34	0.34	0.38	0.52	0.31	0.32	0.35	0.332
Avg	0.36	0.32	0.35	0.39	0.46	0.36	0.33	0.35	0.3323

Figure 6: Comparison of Genetic Algorithm with other models

Average F1-SCORE for 50 videos using Genetic Algorithm is 0.3323

The results show our model perform better the Random Sampling and Archetypal Analysis 1.

Chapter 6

Conclusion and Future Work

6.1 Conclusion

In this paper, we developed a video summarising method using genetic algorithm based on Object Of Interest based video summarization. The results show that the proposed method is capable of producing summaries that are both representative and diverse. Our study emphasizes the value of utilizing optimisation algorithms for video summarization and highlights how crucial it is to consider both representativeness and variety when creating summaries. After performing tests on the given TVSum dataset, we have seen an improvement in the F1 scores as compared to the techniques like Random Sampling and Archetypal Analysis 1 given in the CVPR2015 research paper.

6.2 Future Work

Video summarization using genetic algorithm has several potential future research directions. Fitness function mostly determines the performance of genetic algorithm.Future studies could examine the use of various fitness factors to enhance the precision and variety of the generated summaries.We can add various additional parameters for calculating fitness function.We can take a single frame as input depicting object of interest of the user and include those frames in final summary that

are similar to given frame.We can also prioritize frames having text content.

Genetic algorithm can be enhanced by incorporating domain-specific knowledge such as scene detection, object recognition, or audio analysis to improve the quality of the summary. Future studies can look into how deep learning and genetic algorithms can be used to create more effective and efficient video summarising methods.

References

[1] Evlampios Apostolidis, Eleni Adamantidou, Alexandros Metsai, Vasileios Mezaris, and Ioannis Patras. Video summarization using deep neural networks: A survey. *Proceedings of the IEEE*, 109:1838–1863, 11 2021.

[2] Jayanta Basak, Varun Luthra, and Santanu Chaudhury. Video summarization with supervised learning. pages 1–4, 12 2008.

[3] Zakaria Abdelmoiz Dahi and Enrique Alba. The grid-to-neighbourhood relationship in cellular gas: from design to solving complex problems. *Soft Computing*, 24, 03 2020.

[4] Upulie Handalage and Lakshini Kuganandamurthy. Real-time object detection using yolo: A review. 05 2021.

[5] Tao Xiang Kaiyang Zhou, Yu Qiao. Deep reinforcement learning for unsupervised video summarization with diversity-representativeness reward. 18 2017.

[6] Ying Li, Shih-Hung Lee, Chia-Hung Yeh, and C.-C. Jay Kuo. Techniques for movie content analysis and skimming tutorial and overview on video abstraction techniques. *Signal Processing Magazine, IEEE*, 23:79 – 89, 04 2006.

[7] Naveenkumar Mahamkali and Vadivel Ayyasamy. Opencv for computer vision applications. 03 2015.

[8] HS Pransantha. Video segmentation summarization using modified genetic algorithm. 2018.

[9] Bashir Sadiq, Bilyamin Muhammad, Muhammad Abdullahi, Gabriel Onuh, Abdulhakeem Ali, and Adeogun Babatunde. Keyframe extraction techniques: A review. volume 19, pages 54–60, 01 2020.

[10] Michael Smith and Takeo Kanade. Video skimming and characterization through the combination of image and language understanding. pages 61 – 70, 02 1998.

[11] Yale Song, Jordi Vallmitjana, Amanda Stent, and Alejandro Jaimes. Tvsum: Summarizing web videos using titles. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5179–5187, 2015.

[12] Manasa Srinivas, Manohara M M, and Radhika Pai. An improved algorithm for video summarization – a rank based approach. volume 89, pages 812–819, 12 2016.

[13] Xue Yang and Zhicheng Wei. A video summarization using a genetic algorithm. volume 7, pages 341–350, 04 2012.