A Method for Static Video Summarization for Reducing Power Consumption and Its Impact on Bandwidth

طريقة تلخيص الفيديو الثابت لخفض استهلاك الطاقة ومدى تأثيرها على معدل نقل البيانات

Prepared by

Mohammed Hamid Al-Kubaisi

Supervisor

Dr. Mohammed A. F. Al-Husainy

A Thesis Submitted in Partial Fulfillment of the Requirements for the Master Degree in Computer Science

Department of Computer Science

Faculty of Information Technology

Middle East University

Amman, Jordan

August, 2016

AUTHORIZATION STATEMENT

I, Mohammed Hamid Al-Kubaisi, authorize the Middle East University to provide hard copies or soft copies of my thesis to libraries, institutions or individuals upon their request.

Name: Mohammed Hamid Al-Kubaisi

Signature: محمد الكبيسي

Date: 16 / 11 / 2016

أقرار تفويض

أنا محمد حامد الكبيسي ،أفوض جامعة الشرق الأوسط للدراسات العليا بتزويد نسخ من رسالتي ورقياً وإلكترونياً للمكتبات أو المنظمات أو الهيئات والمؤسسات المعنية بالأبحاث والدراسات العليا عند طلبها.

الاسم : محمد حامد الكبيسي

التوقيع: محمد الكبيسي

التاريخ: 16 / 11 / 2016

Examination Committee Decision

This thesis "A Method for Static Video Summarization for Reducing Power Consumption and Its Impact on Bandwidth" was discussed and certified on 3 / 8 /2016.

Thesis committee		**Signature**
Dr. Mohammed A. F. Al-Husainy	Supervisor	
Dr. Sharefa Murad	Chairman	
Dr. Mohammed Ababneh	Member	

ACKNOWLEDGMENT

I knew from the beginning that pursuing graduate study is a difficult and challenging task. Throughout this long journey, I have gained a lot by learning to persevere despite hardship. I am grateful for all of the support and contribution along this journey. I would never have successfully completed this thesis without the assistance of numerous people who I am indebted to. I am forever indebted to my family who supported me during this academic journey and my whole life; they had more faith in me than could ever be justified by logical argument.

I would like to express my sincerest appreciation to my supervisor Dr. Mohammad Al-Husainy whose thoughtful consideration and guidance has been invaluable. Without his guidance, support and inspiration during the most critical period of my Master degree journey, I would not have been able to accomplish this study. To him, I wish to say 'You are a wonderful Supervisor, Doctor and Researcher!'

Many sincere thanks also go to the Information Technology Faculty members at the Middle East University for their insightful conversations and hard questions, thank you for teaching me how to be a dedicated researcher.

Finally, million thanks go to my fellow colleagues for their support and encouragement

Dedication

{وقُلْ ربِ زِدْنِي عِلْمًا} [طه] 111

I dedicate this thesis to my mother who standing beside me and my Father God's mercy. I hope to reach my research into the world to benefit from it and be ongoing charity to dear my father

TABLE OF CONTENTS

Title I

Authorization Statement.. II

إقرار تفويض.. III

Examination Committee Decision IV

Acknowledgment V

Dedication VI

Table of Contents VII

List of Tables IX

List of Figures X

List of Abbreviations.. XI

Abstract.. XII

الملخص... XIV

1.1 Introduction..2

1.2 Video Summarization 2

1.2.1 Static Video Summarization.. 4

1.2.2 Dynamic Video Summarization 5

1.2.3 Key Frame Extraction 5

1.3Problem Statement 6

1.4Aims and Objectives 6

1.5Motivation.. 7

1.6Research Importance 7

1.7Research Questions 7

1.8Scope of the Study 8

1.9Thesis Outline 8

2.1 Background 10
2.2 Video Summarization 10
2.3 Types and Techniques of Video Summarization 13
2.3.1 Dynamic Video Summarization 13
2.3.2 Static Video Summarization 17
2.3.2.4 Clustering Techniques Video Summarization 20
2.3 Related Works 22
2.4 Application of Video Summarization 25
2.4.1 Case Study of a Soccer Game 27
2.5 Summary 29
3.1 Introduction 32
3.2 Methodology and the Proposed Work 34
3.2.1 Stage One: Splitting Process 34
3.2.2 Stage Two: Video Processing 34
3.2.3 Stage Three: Sound Processing 35
3.2.4 Stage Four: Merging 35
3.3 Power Consumption and Bandwidth Model 44
4.1 Introduction 47
4.2 Data Set 47
4.3 Experimental Setup 47
4.4 Experiment 48
4.5Evaluation 59
5.1 Conclusion 66
5.2 Future Work 67
Reference 69

LIST OF TABLE

Table (3-1): Frame Processing.. 37
Table (3-2): Ratio Finding 38
Table (3-3): Result of Frame.. 40
Table (3-4): Collecting of Average Frames 41
Table (3-5): Sound Processing.. 41
Table (4-1): Power Consumption and Bandwidth before and after Summarization 54
Table (4-2): Time before and after Summarization.. 56

LIST OF FIGURES

Figure (1-1): generic approach for video summarization adapted 4
Figure (2-1): Scheme of video abstraction.. 12
Figure (2-2): Video skimmingtechnology process 15
Figure (2-3): Video Summary using dynamic video summarization technique 16
Figure (2-4): Hierarchical structure of video sequence 17
Figure (2-5): Keyframe summarization technique steps.. 18
Figure (2-6): Clustering Techniques Video Summerization.. 20
Figure (2-6): VGRAPH Approach process.. 21
Figure (2-7): VSUMM video summarization process 24
Figure (2-8): Related literature proposed video summarization techniques and procedures........ 25
Figure (3-1): Diagram of the proposed method frame work... 33
Figure(3-2): Video content…………………………………………………...……………36
Figure (3-3): Flowchart of the proposed model 43
Figure (3-4): Flowchart of the power model…………………………………………………..44
Figure (4-1): PSD for test1 49
Figure (4-2): power for Test1 50
Figure (4-3): power after summarization for Test1 51
Figure (4-4): PSD for Test2 52
Figure (4-5): power for Test2 53
Figure (4-6): power after summarization for Test2 54
Figure (4-7): BW1 and BW2……………………………………………………………….55
Figure (4-8): P1 and P2………………………………………………………………………..55
Figure (4-9): Power and bandwidth relation…………………………………………………..57
Figure (4-10): audio signal after filtration and removing partitions ……………………………57
Figure (4-11): spectrogram of the audio data…………………………………………………...58
Figure (4-12): Summarized video frames………………………………………………………58
Figure (4- 13): PSD for Jadhav and Jadhav……………..……………………….........…….60
Figure (4-14): power for Jadhav and Jadhav ……………………...……..……….…………...61
Figure (4-15): power after SUM for Jadhav and Jadhav………………………………………61
Figure (4-16) Video Summary result……………………………………………………………63
Figure (4-17): Video Summary result using proposed method…………………………………64

LIST OF ABBREVIATION

MPEG	Moving Picture Experts Group
VSUHCM	Video Summarization using Higher Order Color Moments
DOI	degree of interest
TWF	Time warp football
CDL	capture-to-display latency
PSNR	Peak Signal Noise Ratio
PSD	Power Spectral Density

A Method for Static Video Summarization for Reducing Power Consumption and Its Impact on Bandwidth

By

Mohammed Hamid Al-Kubaisi

Supervisor: Dr.

Mohammed A. F. Al-Husainy

Abstract

Nowadays, the rapid development of digital video and the need for faster browsing of large amount of data and content indexing, has led to the need for effective and advanced techniques for analysis and video retrieval, since multimedia sources have made browsing, delivery of contents (video) and video retrieval very slow. Hence, video summarization proposes various ways for faster browsing of large amount of data and also for content indexing that reduces cost (bandwidth), memory and requirement energy.

In this thesis, video summarization approaches are discussed, through considering the power consumption and bandwidth for different videos test data through using the main parameters that measures the time, power and bandwidth before and after summarization, in addition to the power spectral density for each test. Research done in this domain is discussed. We conclude that there is a broad perspective for further research in this field.

The proposed strategy is represented by picking diverse test recordings that contain video and sound information, the proposed technique split them into two processing and apply video rundown by dropping particular frame in the video as per multi-level calculation that deduct unimportant edges in unique video and sound sign after rundown is arranged by expelling particular purposes of the example sound sign so as to smooth the general video.

The outcomes demonstrate the connection between power utilization and transmission bandwidth, where diminishing power utilization contrasted with unique video, while there is expanding in allotted data transmission at certain video test. The general power decrease around 80% of unique video size and the impact of transmission bandwidth expanded around 40% overall identified with the transfer speed of unique video.

Keywords: Digital Video, Video Summarization, Frame, Power Consumption, Bandwidth, Sound Information.

طريقة تلخيص الفيديو الثابت لخفض استهلاك الطاقة و مدى تاثيرها على معدل نقل البيانات

اعداد

محمد حامد الكبيسي

إشراف

د. محمد عباس فاضل الحسيني

الملخص

في الوقت الحاضر، ونتيجه للتطور السريع للفيديو الرقمي والاحتياجات لسرعة التصفح بين كمية كبيرة من البيانات وفهرسة المحتوى، استدعت الحاجة لتقنيات فعالة ومتقدمة لتحليل واسترجاع الفيديو. وعند زيادة الحاجة لاستخدام وسائل الوسائط المتعددة ووسائل التواصل الاجتماعي الحاج ملحه لتبادل المعلومات المرئية مثل الفيديوهات، فقد استدعت الحاجة لإيجاد وسائل لتقليص حجم الفيديوهات المتداولة وذلك لزيادة سرعه التصفح وبالتالي توفير الوقت والجهد والطاقة المستخدمة.

في هذه الأطروحة، تم طرح أحد الطرق المستخدمة لتلخيص وتقليص حجم الفيديوهات وأثر ذلك على كل من الطاقة المستهلكة والنطاق الترددي وملاحظه الفرق قبل وبعد التعديل حيث أن ما تم ملاحظته هو تقليل حجم الطاقة المستهلكة بنسبه 80% وزيادة في عرض النطاق الترددي بنسبه 40%، وذلك من خلال اخذ عينات من الفيديوهات المحتوية على مقاطع صوتيه أيضا. ما نعمل على تطبيقه هو فصل مقاطع الفيديو عن الصوت ومن ثم القيام ببعض العمليات التعديلية على كلا المقطعين وبتنفيذ الخوارزميات المذكورة في الفصل الثالث يتم تقليص حجم الفيديو ومن ثم دمج المقطع الصوتي المناسب حسب المقاطع المكوّنة للفيديو الجديد.

وأخيرا فقد تمت مناقشه العديد من الأبحاث التي تدور حول نفس المجال وما تم استنتاجه هو أن الحاجة ما زالت ملحه لتطوير المزيد.

الكلمات المفتاحية: الفيديو الرقمي، تلخيص الفيديو، الإطار، استهلاك الطاقة، عرض النطاق الترددي، معلومات. الصوت

Chapter One

Introduction

1.1 Introduction

Social media has been attributed as one of the most important internet development. Social media is one of the more applications that are used in the world. it includes communication means between individuals sharing resources such as videos and photos.

Video is an advanced multimedia that has been enabled by the availability of internet in the communication field. This multimedia plays an important role in video development techniques. Video can be defined as a set of sequence frames and contains set of images and sounds that makes up the video data.

Increasing in the use of internet videos has brought about the need to summarize the videos to smaller size to make video available for using in multiple networks easily at low cost. Summarization helps the user to determine if the video is worth watching or not. Advanced technologies can be used in video summarization, especially in areas like sport videos, movies and other video clips (Kumar, Kumar, and Majumdar, 2012).

1.2 Video Summarization

Video summarization is a technique used in creating shorter videos from original long videos. There are two major methods of video summarization, these are static and dynamic summarization. The summary extracted using these methods are Skimming (where moving images are summarized) or Key frames (Where stationary images are summarized).

Technique of video summarization has several advantages, first, the user is able to watch shorter videos from the original video which may be hard to read full video where a summary of it is what is required to understand. Again, when one is using mobile phones, the video summary helps save memory as well as increasing the battery life in downloading as well as reducing the cost to download a video especially because short videos requires minimal amount of data. Again, the video summarization methods allow one to browse faster especially for indexing contents given that many people are online streaming information such as soccer, movie, and downloading music.

When it comes to navigation and browsing, a good abstract of a video allows and enables the user to gather and gain information maximally about the sequence of the target video in particular constraints of time; this implies that the user can acquire enough information within the minimum time possible. Videos that contains automatically generated summaries enables and supports the user to navigate and go through extensive archives of videos and make more efficient decisions when it comes to the selection, consumption, and sharing the videos or whether to delete such videos.

Cahuina and Chaves (2013) illustrated the four general steps of video summarization. The first step is the video segmentation followed by the feature extraction, then the redundancy deduction on features and finally the summary of a video is generated. In the figure 1-1 below, summarization of video using the generic approach has been demonstrated.

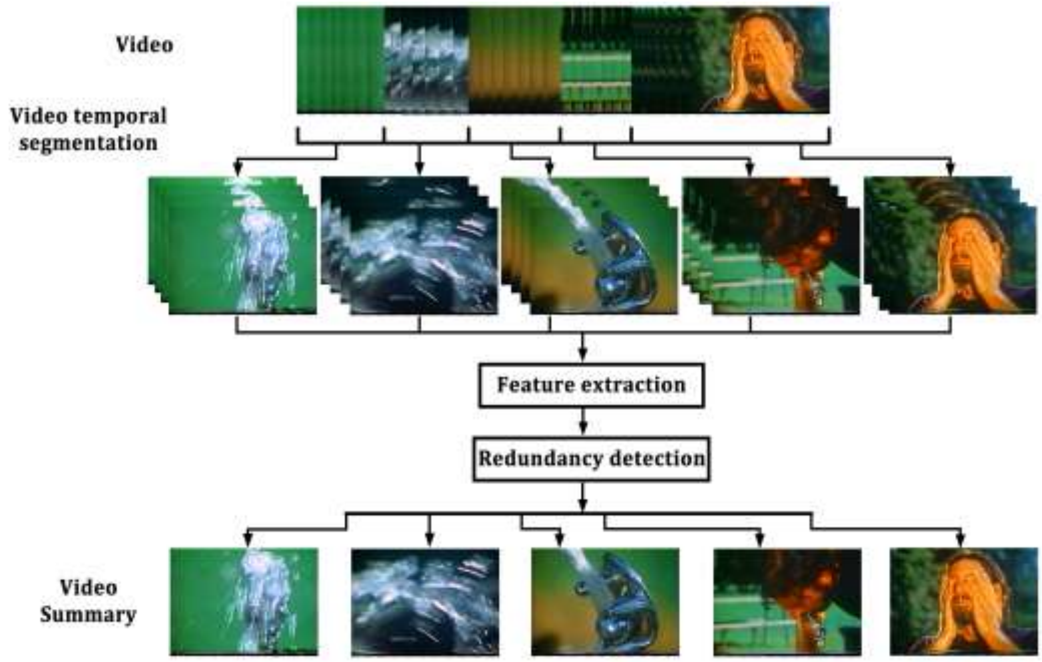


Figure (1-1): generic approach for video summarization adapted from Cahuina and Chavez (2013).

1.2.1 Static Video Summarization

Static video summarization is commonly known as the Reference Frame (R-Frame) also referred to as still image abstract or static storyboard. Three major methods are used to classify the static summarization. They include shot segmentation, scheme segmentation and video based on sampling. The video takes key frames uniformly or randomly. Key frame is viewed the process of video content management which help chose one frame or part of multiple frames in order to generate a video summary (Khan and Pawar, 2015).

Scene segmentation is classification of the key frames is done by use of scene detection, included in the scenes are all parts that have a semantic link in the original video or in the same time or in the same space. Shot segmentation is where the adopted key frames are extracted to video contents. The first image is extracted as shot key frame or the shot's first and last frame. Finally, clustering technique is the notion is clustering together same shots/frames, followed by extraction of the same frame per cluster as a key frame. The techniques differ in terms of features as luminance, color histogram, motion vector as well as clustering algorithm as k- means, hierarchical.

1.2.2 Dynamic Video Summarization

Dynamic video summarization also commonly referred to as the video skim is used in generation of short videos from the original ones. A technique used to get dynamic video application is known as the skimming semantic analysis(Khan et al., 2015). In segmentation of videos into scenarios with semantical meaning where shorter videos are acquired, the video relaying on 2-domensional histogram entropy of pixel of image is adapted. The common types of video summarization under this technique are the motion model and the Singular Value Decomposition (SVD).In comparison with the static method, skimming supports object recognition in the video content, the representativeness of the object is adequate to replace the original contents of the video.

1.2.3 Key Frame Extraction

In key frame extraction, a key frame is received and produced being same to the original image which is a sub-short and should be dissimilar to the framework of the outside video at similar time. The key frame of a video is viewed as one of the frames of a video providing the best summary of the video content. The key extraction has played a critical role in video summarization where most of the common techniques used in the method of key frame extraction are based on macro-block statistical characteristics of MPEG video streaming, based on Shot activity and finally based on motion analysis.

1.3 Problem Statement

Nowadays, social media is one of the most used applications on a large scale in our daily lives. Video is considered one of the most objects used in social media. Most people get the videos through social networks and this video may not contain important information or be repeated. However, the user is left with the only option is to download the video. This leads to add more and more of the costs to users because the video requires a large bandwidth to download video or view it and a large space to store it. Thus, it is important to be a way to enable users to watch videos in a way that helps to reduce costs. Video summarization technique is a proposed solution to resolve different points such as power consumption and additional costs.

1.4 Aims and Objectives

The Objective of thesis is to create a short summary that can still convey the original story , with a good power consumption decrease , in addition to increase of the bandwidth.

To achieve the above purpose, we go through the following:

- Conduct process on video frames to eliminate any bulky or duplicate data (frames) to choose the best ones.
- Conduct process on audio signals so that it suited with the new video.
- Apply the suitable algorithm to obtain a good summarized video.

1.5 Motivation

The increasing for the use of modern mobile devices, increasing multimedia content and Due to possess these devices modest sizes and weights. This means it has limited computing, storage and battery resources. Moreover, the multimedia applications developments and increase in content has brought about significant challenges in research field in mobile with an objective of reducing the cost.

Therefore, in this work we focus on achieving the video summarization in the mobile device without significantly compromising on the quality of the video. Researchers have proposed several schemes in recent years for video to optimize the costs in mobile environment.

1.6 Research Importance

In this work, different parameters can be studied. The following are the main points that provide the importance of the research

- Relation between the power consumption and dropping specific number of frames.
- Relation between the effect of required bandwidth before and after video summarization.

1.7 Research Questions

- How much power can be saved in video summarization technique?
- How much power consumption effect on bandwidth?
- How to split and combine video and audio signals?

- How can we remove un-necessary audio signals between frames?

1.8 Scope of the Study

- Specific videos will be tested , one of them will be with audio signal as a case study
- Shannon capacity will be used to build mathematical model between power and bandwidth.
- Total capacity of this work is assumed to be constant. (10M/sec/Hz).

1.9 Thesis Outline

Chapter One: Introduction, and importance of the study, techniques of video summarization, objectives of the thesis and problem statement.

Chapter Two: explains the main concepts of the video summarization and presented some of the literature reviews that related with this thesis.

Chapter Three: presents the proposed methodology that has been followed in practical part of this thesis and represents it in algorithm and flowcharts.

Chapter Four: presents a number of experiments through interfaces and evaluates each experiment by results.

Chapter Five: provides conclusion and suggestions for future works.

Chapter Two

Background and Literature review

2.1 Background

Since the use of the videos in the web has increased over the years, there enormous numbers of studies and the application based on video summarization techniques have been proposed and it has been looked at from multiple perspectives.Even though lots of papers and researches were considering the costs in mobile environment but not specifically concentrating on the relation between the power consumption and dropping specific number of frames and the relation between the effect of required bandwidth before and after video summarization. This will reveals the importance of this thesis where unlike previous approaches, it focuses on How much power can be saved in video summarization technique and bandwidth and how to split and combine video and audio signals and How can we remove un-necessary audio signals between frames.,

In this chapter, video summarization deeply will be discussed, Types and Techniques of Video Summarization (Dynamic Summarization, Static summarization and Clustering techniques summarization) will be focused on, in addition to the related works as well as the application of Video Summarization.

2.2 Video Summarization

Basically, video summarization can be defined as a technique or mechanism for producing a shorter video than the original one and to pocess vidoes that contains redunduncy in order to make them more interesting and valuable. More over, providing users a useful synthetic visual abstract of the sequence of the video which can either be in form of a moving image (video skims) or images (key frames) (Potapov et al., 2014). As mentioned before, the rapid increase in the video content volume downloaded from the web besidesthe effort requirement to download and process videos, reveal the need for efficiently and effectively new technologies that could manage the large amount of data in the video content.

When it comes to navigation and browsing, the user can acquire enough information within the minimum time possible since a good abstract of a video enables the user to gather maximum information for the sequence of the target video shorter time.

Videos that contains automatically generated summaries, enables and supports the user to navigate and go through wide-ranging of videos archives to allow the user to make efficient decisions while selection, consumption, and sharing the videos or to delete such videos (Rehman and Saba, 2014).

Moreover, with this technology, the user can use the end product to share, digest or even enjoy the contents of the summarized video which in turn, saves and improves the storage of the video contents efficiently, saves the bandwidth when it comes to downloading the video besides the viewing time for the summary is saved. There are numerous domains touched by the development of the techniques of video summarization; such as e-learning, home videos, news, sports, movies among other areas (Furini et al., 2010).

There are three phases of the process of video abstraction. The first step is the analysis of video information, followed by a selection of the clip that is meaningful and Finally being output synthesis (Zhang et al., 2015).

To do the analysis of the video information, it is crucial that relevant features, patterns or structures be identified in audio, textual and visual components like closed captions. In figure (2-1), the phases of the video abstraction process are presented.

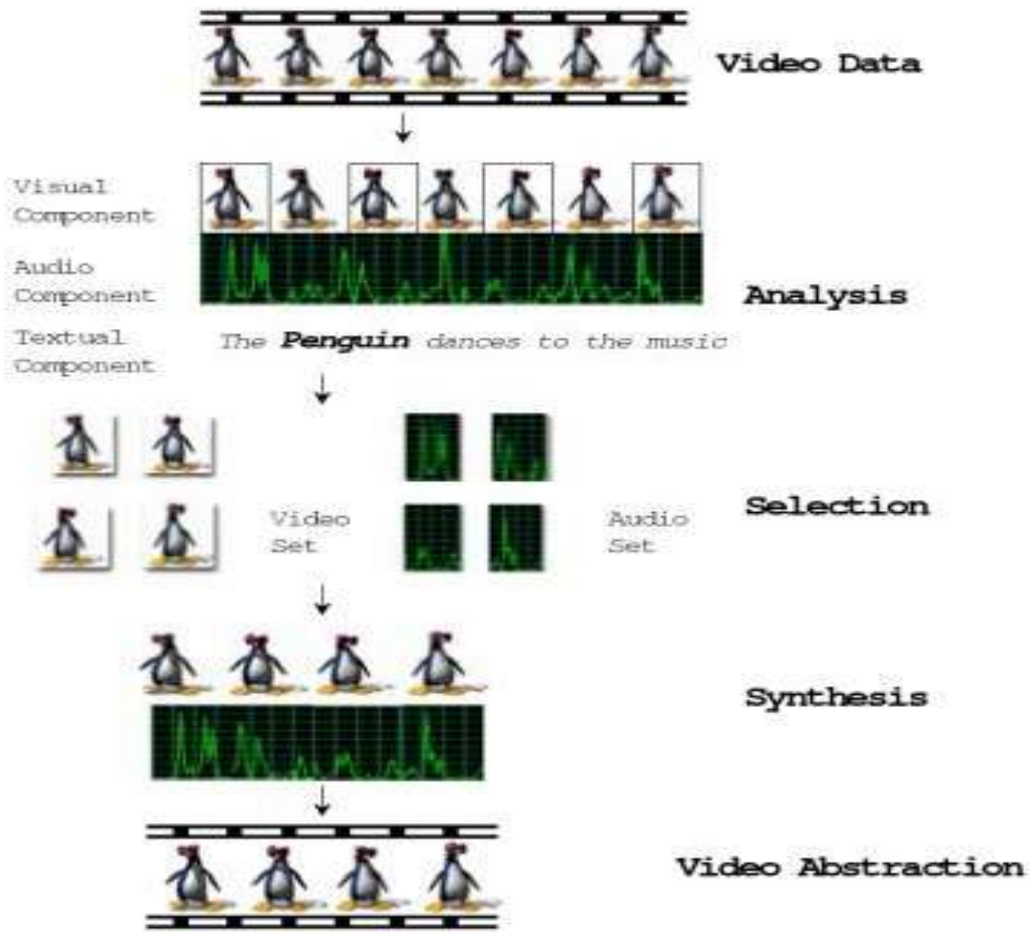


Figure (2-1): Scheme of video abstraction (Zhang et al., 2015)

There are two modes in which video summarization can be represented; these are the dynamic video skimming and Storyboard (static video summary) (Saba and Altameem, 2013). Static video summary usually talks about the storyboard that used in summarizing video in a static imagery form; this method allows keyframes to be extracted from the original video shots. After that, the selected key frames are blended or arranged in twodimension space. In comparison, dynamic video summary helps choose the most significant and relevant small portions that are dynamic; these are called video skims and have both video and audio and are used in the generation of the summary of the original video (Sigari et al., 2015).

In the following section, the video summarization techniques and types have been discussed in depth.

2.3 Types and Techniques of Video Summarization

In this section, three main video summarization techniques will be discussed, these are the Dynamic Summarization, Static summarization and Clustering techniques summarization (Lee et al., 2012).

2.3.1 Dynamic Video Summarization

In dynamic summarization of videos, the extracts of the audio and visual segments are taken from the original video based on the notion of video skimming to get shorter videos containing important scenes and data of the original video (Knox et al., 2014).

In this case, the user can be able to receive an abstract view of the entire video, which is known as the video story. In this skimming technique, a segment of the original video is taken containing the audio and the video summary part of the original one. The common types of video summarization under this technique are the motion model and the Singular Value Decomposition (SVD) (Mayberry et al., 2014). The method of semantic analysis can also be applied in this skimming technique. While most of the skimming techniques are based on the visual information, other methods/technologies have been viewed to use and implement both linguistic and audio information.

In the skimming video abstraction, the movie video schemes are presented. The approach tries to get the contents of a video from the progress of the human semantic understanding and overall story (Lu et al., 2013).

In the beginning, the two-dimensional histogram property is implemented to enable segmentation of the video into shots. After that, the general rules of common techniques and special scenario of production of video are applied/used to allow the progress of a story to be grasped regarding the progress degree between the scenarios to the overall story (Lee et al., 2012). The Process of video skimming technique is shown in figure 2-2.

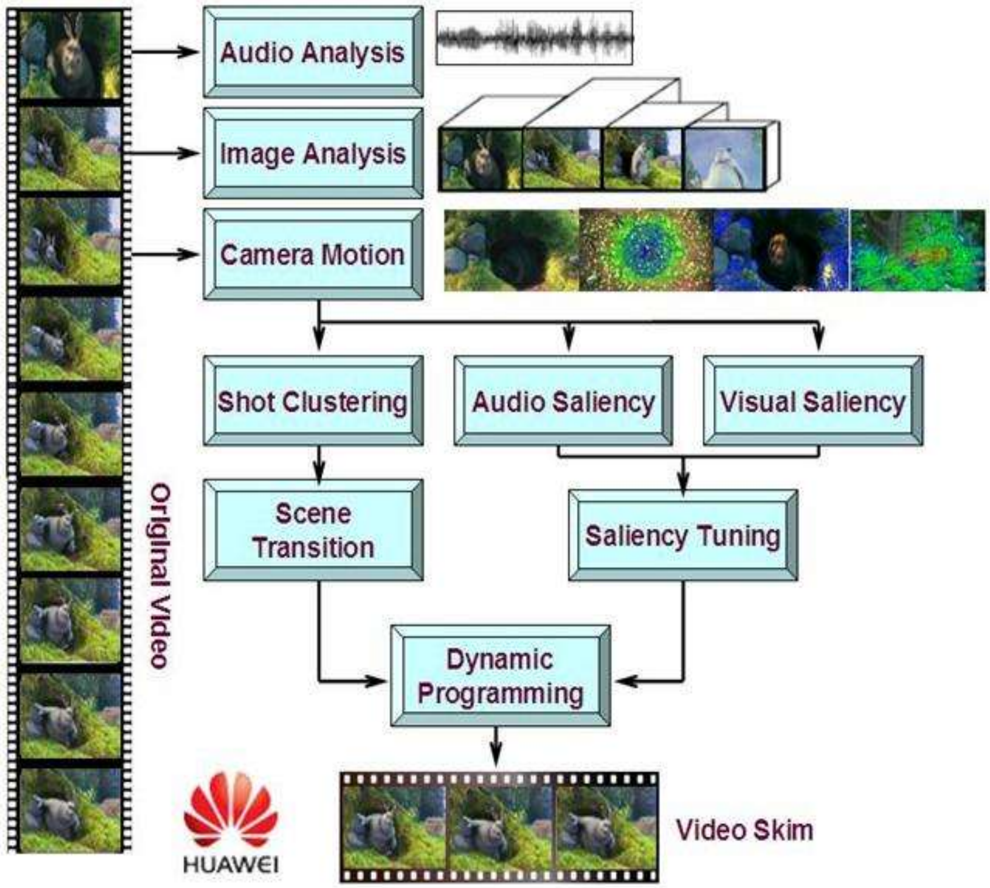


Figure (2-2): Video skimmingtechnology process (Kokaram, 2013)

To make the skimming approach successful multiple information which includes transcript, speech , sound and video image analysis should be used from various sources. A good example is automatic skims of videos of the newscast, it should be considered that the videos consist of textual transcripts (Sigari et al., 2015). First, abstracts of the texts are acquired, this can be done using the skimming technique of classical text skimming, after which the corresponding videos to this textare acquired from the original video. By so doing, skim videos are obtained representing a short original video synopsis. The goal of video skimming in the example above is to acquire and integrate the understanding of image and language from the original video which is done by ensuring that the original information is extracted (Knox et al., 2914). This information consists of audio keywords, specific relevant objects and corresponding video structures as demonstrated in figure(2-3) below consisting of a video skim.

Figure (2-3): Video Summary using dynamic video summarization technique (Saba and Atameem,2013)

In comparison with the static method, skimming supports object recognition in the video content, the representativeness of the object is adequate to replace the original contents of the video.

2.3.2 Static Video Summarization

The static technique which can be also called static storyboard or still-image abstract of video summarization, also commonly referred to as the R-frame is a method used to summarize images from the video.

The methods are involved with the extraction of key frames through pre-sampling the original video sequence randomly or uniformly. Extraction of the key frame is essentialfor the management of video content that involves selecting one or more frames to represent the original video content and use in generation of video contents (Jadhav and Jadhav, 2015). A hierarchical structure of sequence of a video is shown in figure 2-4 below

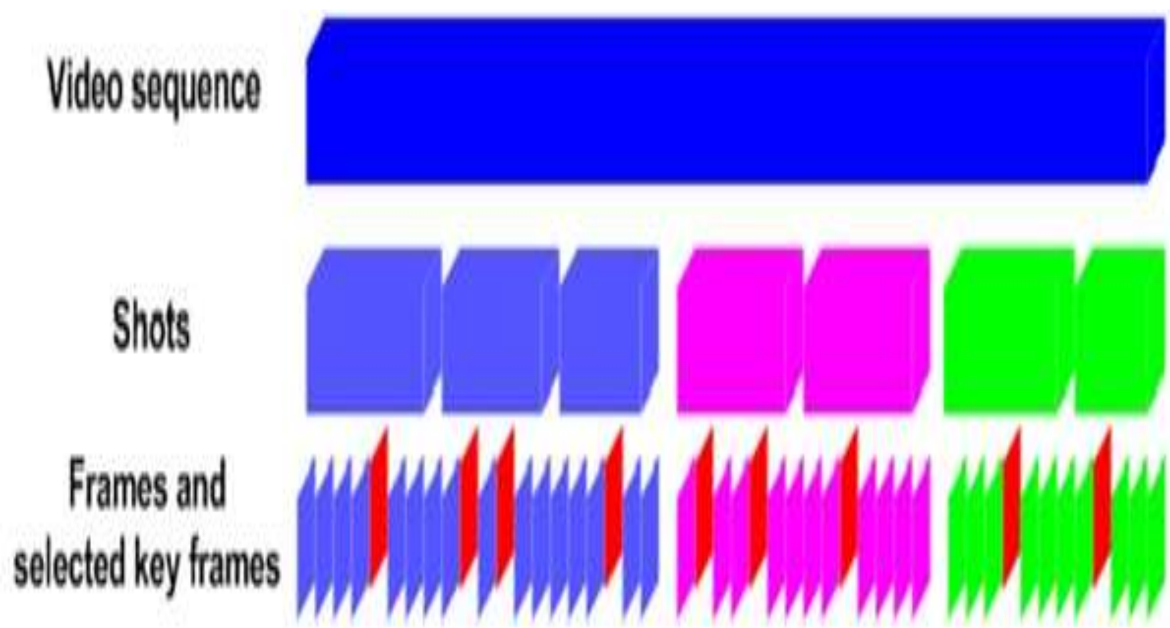


Figure (2-4): Hierarchical structure of video sequence (Jadhay&Jadhay, 2015).

The disadvantage of using this method is that all shots appear with the same level of importance to the user, therefore that summarized output video appears bulky especially for the long videos.

The main steps of key frame video summarization technique (static summarization) are (Lu et al., 2013):

- First, extract the video frame from the frame sequence of the original video.
- The second step is to group the video frameon the basis of the various clustering algorithm, where shot detection is needed from the content deduction.
- Finally, the key frames are selected

Figure (2-5) shows the entire process of the Keyframe summarization.

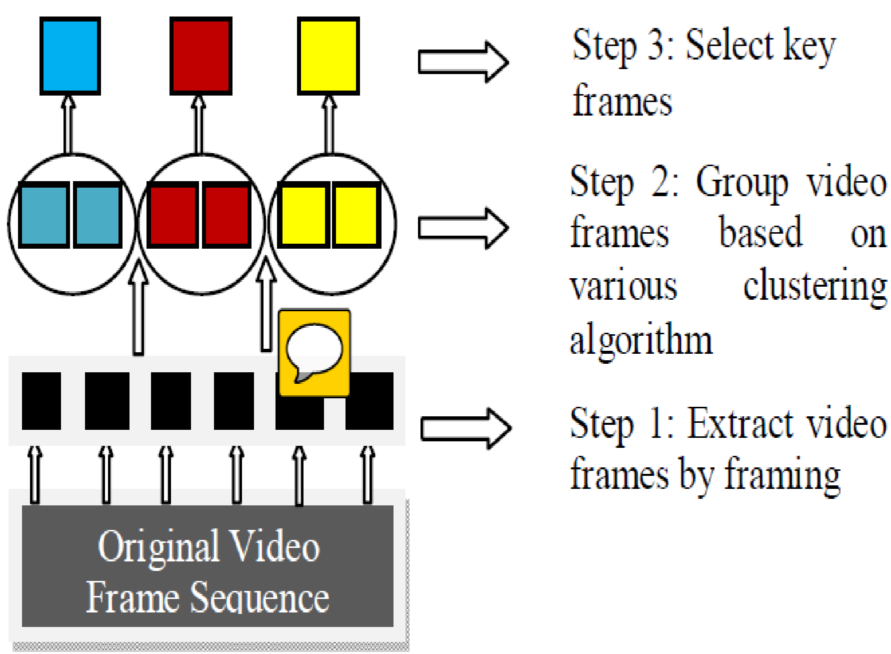


Figure (2-5): Keyframe summarization technique steps (Almeida et al., 2012).

This technique is classified into three various ways (Gasic et al., 2013).:

1. Classification on the basis of sampling,
2. Classification on the basis of scene segmentation or
3. Classification on the basis of shot segmentation.

2.3.2.1 Sampling Classification

Mainly in this classification, the keyframes which have similar contents is redundant since they are chosen randomly and uniformly under-sampling, the video content is not considered and the produced summary is not representative of all the original video parts. (Almeida et al., 2013).

2.3.2.2 Scene Segmentation Classification

In this second method of classification, classification of the key frames is done by using scene detection, all parts are included in the scenes that have a semantic link in the original video or in the same time or in the same space (Almeida et al., 2013). The only disadvantage with this method is that the method does not consider the frame sequential position.

2.3.2.3 Shot Segmentation Classification

This method of key frame summarization of a video the adopted key frames are extracted to video contents. The first image is extracted as shot key frame or the shot's first and last frame. This method of data summarization is most effective in cases where there is small content variation or shots that are stationary (Almeida et al., 2012). However, they do not provide enough representation of shot when there are strong movements.

2.3.2.4 Clustering Techniques Video Summarization

In this method, a summarization is being applied based on which called clusters , on which a pre-sampling is executed to reduce the number of frames based on a specific sampling rate(for example one frame per second) , after that different feature extraction techniques is being applied including color , texture and shape. Here after a clustering algorithm in being applied to get the cluster of the given input frames. Finally a frame is selected from each cluster to make the output summarized video

The step that shows how clustering technique applied is shown here after:

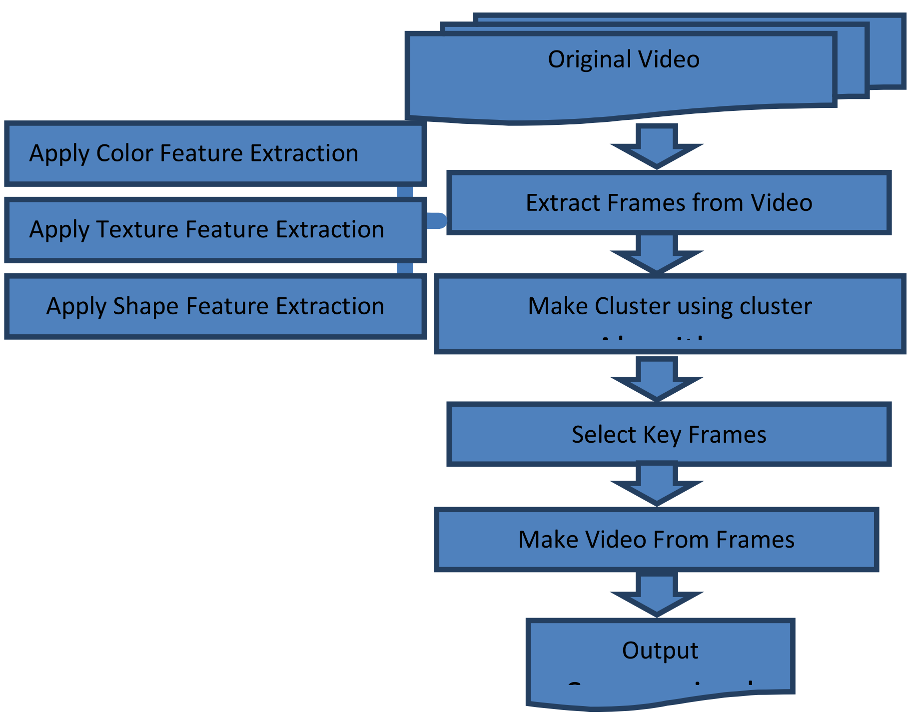


Figure (2-6): Clustering Techniques Video Summerization

2.3.2.5 VGRAPH Approach

The VGRAPH approach is the most commonly used in the generation of static video summary. To start with, pre-sampling of the original video is done; this is the first step. thenthe use of colour frame does the segmentation of pre-sampled video into shots. After that the elimination of noise frames and selection of the representative frames from each shot. Finally extraction of key frames is done using the strategy of the nearest neighbor frame; theclosestneighbour frame is built from the texture feature which is extracted from representative frames of shots (Potion et al., 2014). The steps are demonstrated in figure 2-6 below

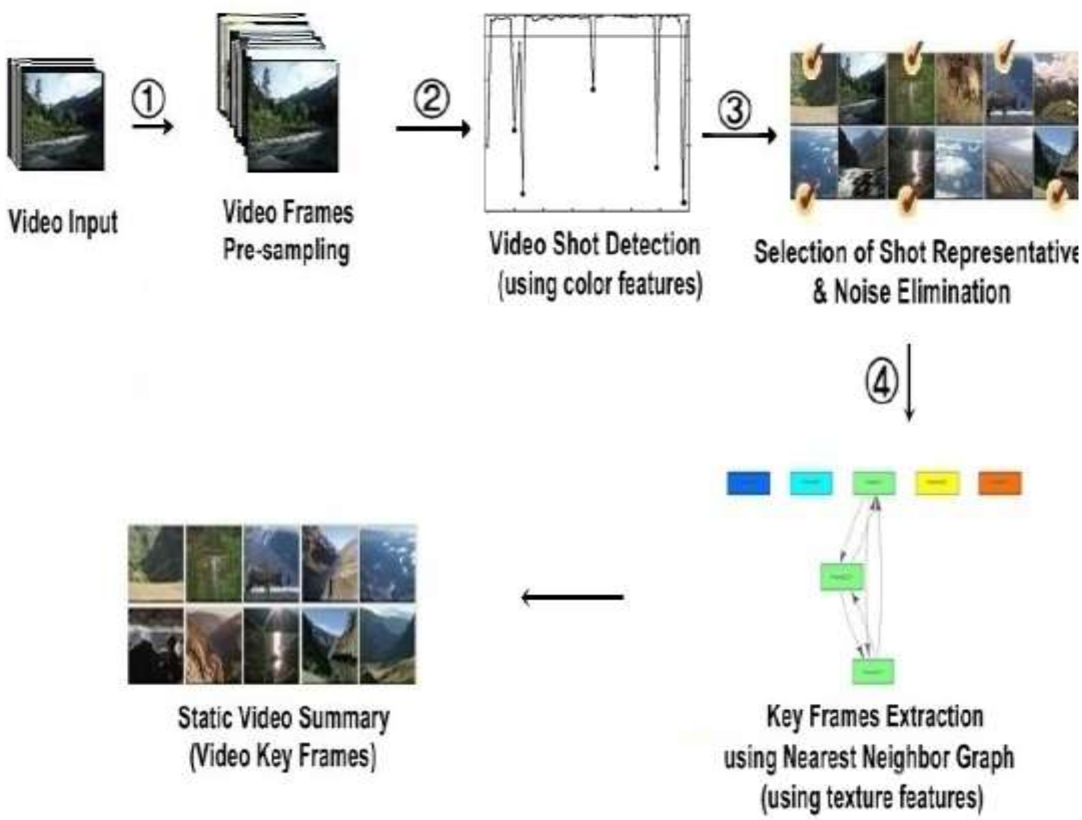


Figure (2-7): VGRAPH Approach process (Potapov et al., 2014).

2.3 Related Works

In the previous works, researchers have done related literature where they have proposed different techniques of video summarization; the related work will focus on such previous research work of video summarization and what some researchers suggested in their work.

One of these researches is done by Mundur et al. (2006) who proposed the use of keyframe technique of video summarization were Delaunay Triangulation (DT) is used. To begin with, the frame content is used as multi-dimensional point data where DT method is used/applied on it to cluster the frame contents (Mundur et al., (2006). In the proposed method by these authors, shot detection is done after the first video frames are extracted. Application of clustering is done on frames of video; this is based on the visual content to acquire key frames.

In another study by Dang et al. (2014, they were proposing the novel approach. In their paper, the researchers used two measure steps, the Quantitative Evaluation and Video Summarization. Firstly, production of video summary is done by using visual features where K-means algorithm is applied. Secondly, there is a process used to get the quality quantitative measures of summary, which involves various users.

Similarly, it was proposed by Zhang et al. (2015) that the most appropriate method of video summarization is the approach of matching variances between two consecutive video frames, which is computed along varying weights. Detection of shot boundaries is done based on an automatic threshold. Finally, the extraction of key frame videos is done by use of the reference frame based technique.

Sigari et al. (2015) proposed the technique which is based on thecolour histogram, one of the lower level features. In their paper, edge detection and colour histogram were the key metrics that were applied for the extraction of the key frame. The primary reason behind this metrics is to ensure that redundancy is eliminated or removed from the frame, to allow for adequate recognition efficiency and less complexity.

As suggested by Schuster and Katsaggelos (2013) in their paper they proposed the recursive multi-dimensional curves algorithm, where this method is used in summarizing the sequence of a video by a small number of keyframes that are small. In the approach, mapping of the video sequence to curve trajectory is done in a high-dimensional space, this is done by use of a feature vector that is carefully crafted to that the significant new information appearance produces high curvature regions or discontinuous information in the trajectory. Local filters of the frame differences are used to detect the key frames of the video sequence (Shuster &Katsaggelos, 2013). Extraction of a set of key frames is done from the concise video sequence, but which contains sample frames extracted from all the sequence shots.

The STILL and moving video storyboard (STIMO) was proposed by Furini et al. (2010)in order to produce on-the-fly storyboards of video and have three main steps. In the first phase, extraction of the feature vector is done from the video frame selected; this is done computing a color histogram in HSV. The second stage is the method of clustering based on the FPF algorithm which is applied (Further-Point-First) (Furini et al., 2010). In order to facilitate for the estimation of the cluster number, the consecutive frame pair wise distance is computed b use of GJD (Generalized Jaccard Distance).The increment of the cluster number increases when the distance between the consecutive frames is greater than the predefined threshold.

The approach of VSUMM was proposed by (Furini et al., 2010)as shown in figure (2-8) below.

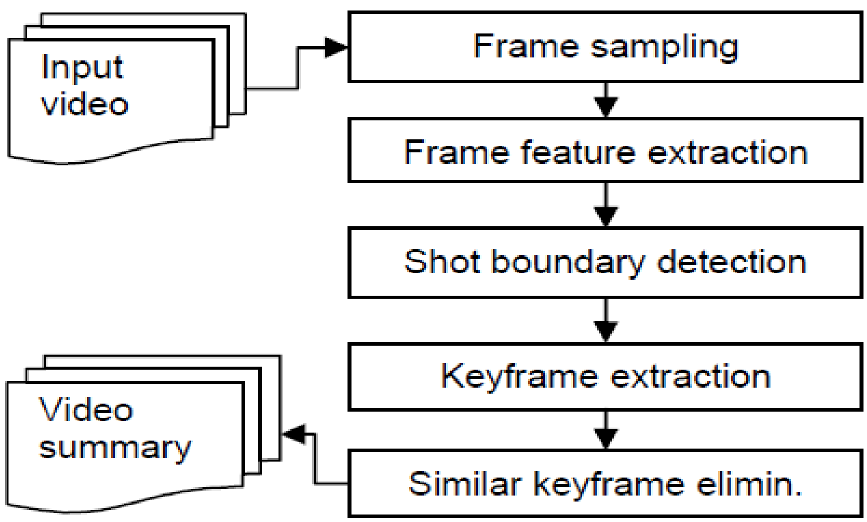
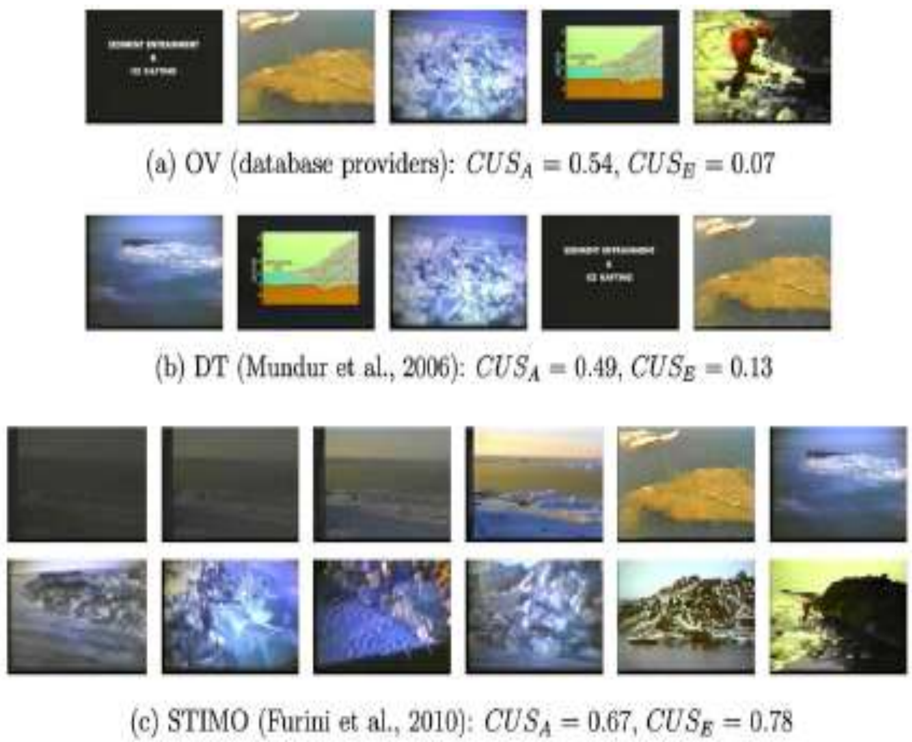


Figure (2-8): VSUMM video summarization process (Furini et al., 2010)

In their first step, frames of the video are pre-sampled through a selection of one frame per second. In the following stage, video frame color features are extracted from hue components, where only HSV color space is used. The third step entails the elimination of the meaningless or insignificant frames of video (Furini et al., 2010). While in the fourth stage k-means algorithm is used to cluster the frames where the estimation of the number of frames is done by computing the pair wise Euclidean distances between the frames of the video to extract key frames from each cluster. In the final stage, there is an occurrence of other extra steps where the comparison between the key frames is done through color histogram with an objective of eliminating key frames that are similar key frames in the video summaries produced.



(a) OV (database providers): $CUS_A = 0.54$, $CUS_E = 0.07$

(b) DT (Mundur et al., 2006): $CUS_A = 0.49$, $CUS_E = 0.13$

(c) STIMO (Furini et al., 2010): $CUS_A = 0.67$, $CUS_E = 0.78$

Figure (2-9): Related literature proposed video summarization techniques and procedures.

2.4 Application of Video Summarization

Many educational and professional fields that is dealing with huge number of videos could be a candidate that may take a good advantage of video summarization applications include image videos, personal videos, sports videos, database management videos, besides the media and television broadcasting company is interested too especially for indexing a large number of videos stored in their e- libraries for internal use.

In this study, the paper will provide application of video summarization for sports videos.

Sports videos can be seen as video contents that consist of similar fascinating events, with the ability to capture the user attention. Many users seem to go for the sports videos which are summarized instead of watching the whole video, which most of the sports games such as soccer goes for even 90 minutes or more. In the full video, there might be many insignificant occurrences or events such as unnecessary playbacks, advertisements or even too many playbacks which may increase the cost to the user without maximizing the benefits of viewing the whole video (Saba and Altameem, 2013). There are several techniques used by the sports websites and broadcasters to edit the sports videos to make them appealing to the users; such include the slow motion scenes and super-imposed text captions with a view togiving preference and discriminate the key soccer events that may be appealing for the users to watch.

Shot boundary detection in sports video is the central part of event detection. There are different methods already in place; these are techniques such as event detection, and frame based segmentation and temporal video segmentation. In shot view classification, detection of various views are contained, these include out of field view, medium view, close up view and long view (Mundur et.al, 2006). Among the techniques that have been proposed, some of them mainly use the frame domain color for the classification of the view. The skin color of the player can be the dominant color in the closed view while in the long view one can pick one color such as green as the background color for the field. For the replays, they are generally played in slow motion where the replays are broadcasted between graphical logos (Rehman and Saba, 2014). By identifying such logos, replays can be detected.

2.4.1 Case Study of a Soccer Game

TV broadcasting companies and media organizations have shown interest in the application video summarization techniques. This is in the effort to increase the efficiency of video streaming for their customers and also reduce the cost and storage associated with the original video which may discourage their viewers from viewing or downloading the video (Almeida et al., 2012). Again, when such media companies broadcast the highlights of sports news and required to have small relevant sections of the soccer game so that they can show a highlight of the game hence this has increased the need for sports video summarization.

In summarizing the soccer news, the dynamic technique of video summarization will be used in this paper for automatic extractions. The method will be based on detection of shot, classification of shot and finite state machines. The main stages that will be discussed in the case study will be four including segmentation of the playfield, detection of a shot by use of DCT-MR (Discrete Cosine Transform Multi-Resolution and finally the word extraction for the soccer video and locating the most appropriate sub-words (Lee et al., 2012. In the subwords, the summaries are presented, where domain knowledge and FSM are used where the rules sets have been definedto have a soccer game that is in the semantic state. Interesting relations between the video semantics and syntactic structures have been explored. At the shot detection step, the various types of shot transmissions are extracted, there are also three major classes that are involved in the shot classification, these are long, medium, and close-up shot, the statistical methods are used to classify the shots.

To detect the most important events in a soccer game such as fault, non-highlights, goals, corners, offside, goal attempts and cards, detection method proposed should be capable of doing this. The most likely method that can be used is the Chow-Liu Tree, which is used in Bayesian Network structure estimation (Lu and Grauman, 2013). This approach has been proposed because it can quickly and appropriately recognize patterns giving approximated results that are good. From the previous applications in video summarization, the Chow-Liu Tree has provided to provide the right or at least better approximation for multivariate probability distribution that is discrete.

To extract the features such as audio and visual features like low-level features and mid-level features, pattern recognition-based technique has proved to be the most appropriate; the events of low and high semantics are easily detected using this approach. In the automatic method, subspace-based methods of data mining are used to extract the features (Mayberry et al., 2014).Given that the method proposed is generic, there is no need for prior knowledge in the process of detection, and this makes it be considered as domain free technique. This method uses the C4.5 decision tree classifier. The MMP (mixture modality project) can also be used in obtaining the high levels of features from medium and low-level features.

Nonetheless, other alternative techniques canbe utilized in soccer video summarization, some of these are dynamic Bayesian Network (DBN) and this is used in capturing the temporal patterns of the features extracted when the soccer event is going on

(Kokaram, 2013). To detect sports highlights, a hybrid approach can be adopted; this helps in integrating the statistics of audio-visual into the model of logicrule based. This method uses the sequence of playback as a sports video semantic unit. The method has been used in sports such as Australian football, basketball as well as football.

To achieve a rapid and accurate events in a video classification, subspace collection method can also be used, where subspace grid method is employed. This technique is prepared to save the intra-modal geometry of the events inside a class that is matching and disentanglement the individual classes (Potapov et al., 2014). A feature vector, with the structure,is able to bring a multi data which is miscellaneous, that can be quickly and efficiently predicted from modalities and identifies that are distinct into unified subspace, this helps to perform recognizing technology. Again, the training phase is completed one time.

2.5 Summary

Escalating the growth of video data in the modern world recently reveal the importance of video summarization that attracted considerable interest from researchers and as a result, various procedures and techniques have been proposed.

This thesis focused on a research that applied static video summarization to achieve a power consumption in addition to high increase in the bandwidth.

In this thesis, chapter two discussed deeply types and techniques of video summarization, in addition to the related works as well as the examples of the applications benefits from video summarization. Then, it focuses on the most important points in this research for using it.

Here after, we present our methodology used in developing the system and the techniques used in details in chapter three. The researcher shows the extracted results of the implemented system in chapter four. Finally provides a conclusion of this work followed by suggestions and motivations for future work in the field of video summarization.

Chapter Three

Methodology and the Proposed Work

3.1 Introduction

A new model has been introduced in this thesis to reduce the bandwidth and power consumption on mobile devices for video. This proposed method aims at giving the user capacity/ability of previewing videos by downloading a video version that is lightweight. Generation of lightweight video is done through creation of small sized video from the original video. The downloading bandwidth cost is reduced by this small size video, also, the power consumption and network traffic and reduced together with ensuing that minimum storage is used. Additionally, the technique is suitably used in on mobile devices that are based on lightweight version.

The method proposed can be demonstrated in four stages: the first stage is splitting the input video file into two parts (sound and video). The second stage is video processing that entails (frame extraction, calculate the difference between each two adjacent frames, determine the time to take frame , find block/key frame, calculate the difference between block frames and selection of the desired block frame on the basis of threshold value. Third stage is sound processing that entails (analysis of sound data and getting the sound data segment desired relative to the block frame selected). The final stage is the merging process to construct a new video summarized. The diagram below (3-1) shows the overall framework of the proposed method.

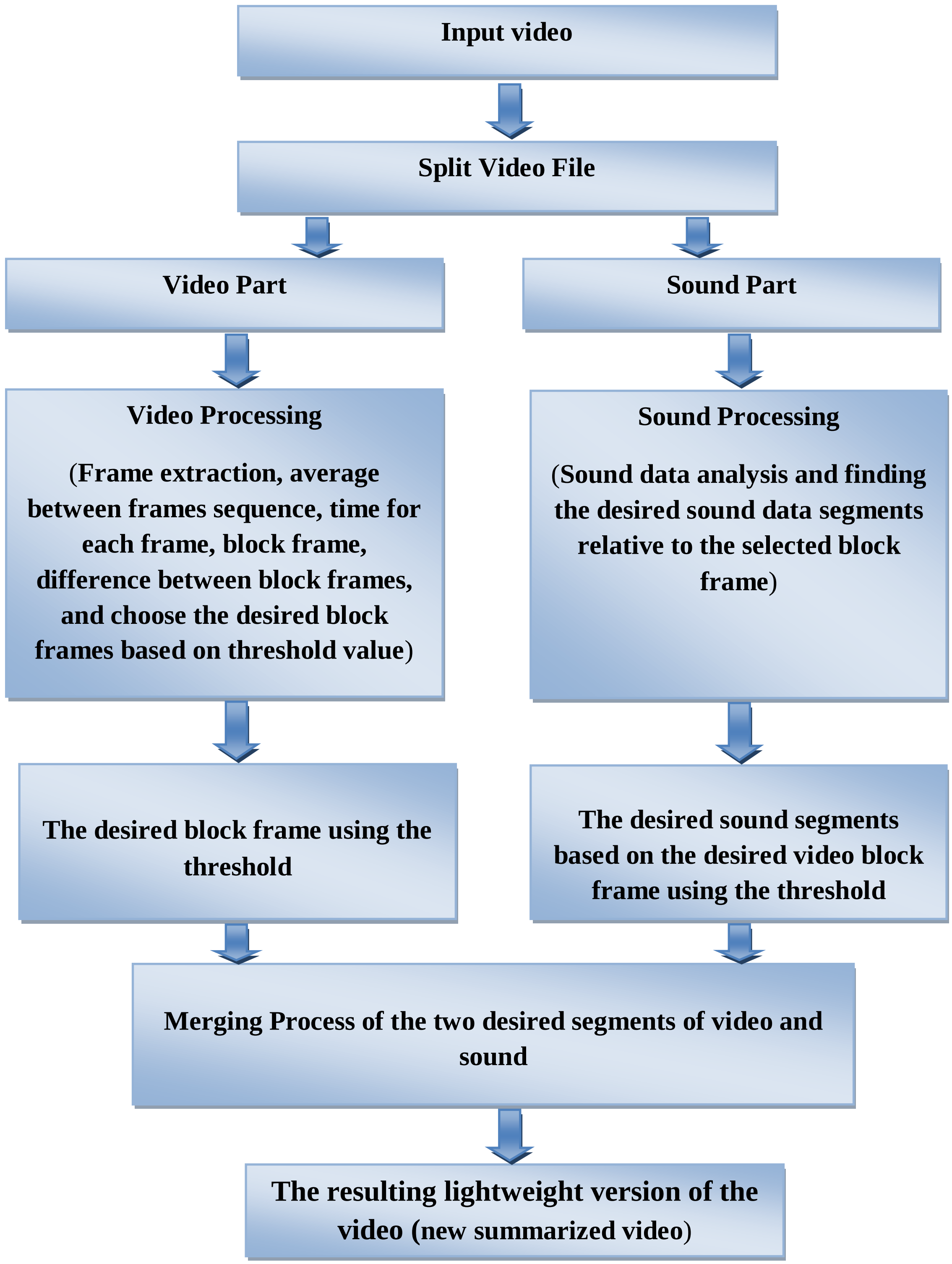


Figure (3-1): Diagram of the proposed method framework

3.2 Methodology and the Proposed Work

The model will give the user the ability to preview video by downloading a lightweight version of the video through creating a new small size video from the source video. The small size video will minimize the network traffic and communications bandwidth cost for downloading and keeping the storage at minimum requirements. The main stages for this study are listed below:

3.2.1 Stage One: Splitting Process

1. Split the video file into two parts:
   - Video part (sequence of frames)
   - Sound part (sound data (wave))
2. Process each part of video file separately as in stage two and three.

3.2.2 Stage Two: Video Processing

1. Extraction the video frames
   - **(f1, f2, … fn).**
2. Calculate the difference between each two adjacent frames.
   - (**i.e., DF1(Difference Frame)=f1-f2 , DF2=f2-f3 , DF3=f3-f4 , … DFn)**……(1)
3. Find the highest percentage differences between the frames and rely on the most frequent ratio.
   - **(MF_{Ratio} =MostFreq($DR_{i,i+1}$))** …………………………….(2)
4. Find block/key frame correspond to each set of frames depend on step (3).
5. Calculate the difference between block frames.

- **(i.e., DB1(Difference Block)= B1-B2 , DB2= B2-B3 , …DBn)**…………….(3)

6. Negligence the blocks depending on the percentage changes between the blocks.
    - **(It will be seen in the example below).**
7. Select the desired block frames depend on predetermined threshold value.
8. Return the blocks to their original frames.

3.2.3 Stage Three: Sound Processing

1. Analysis the sound data to determine the cutting point in the sound wave.
2. Select the desired sound data segments corresponding to the selected block frames in stage two.
3. Cut some of the voices that do not indicate the meaning in speech.
    - **(By searching on zero crossing of audio signal and delete that range).**

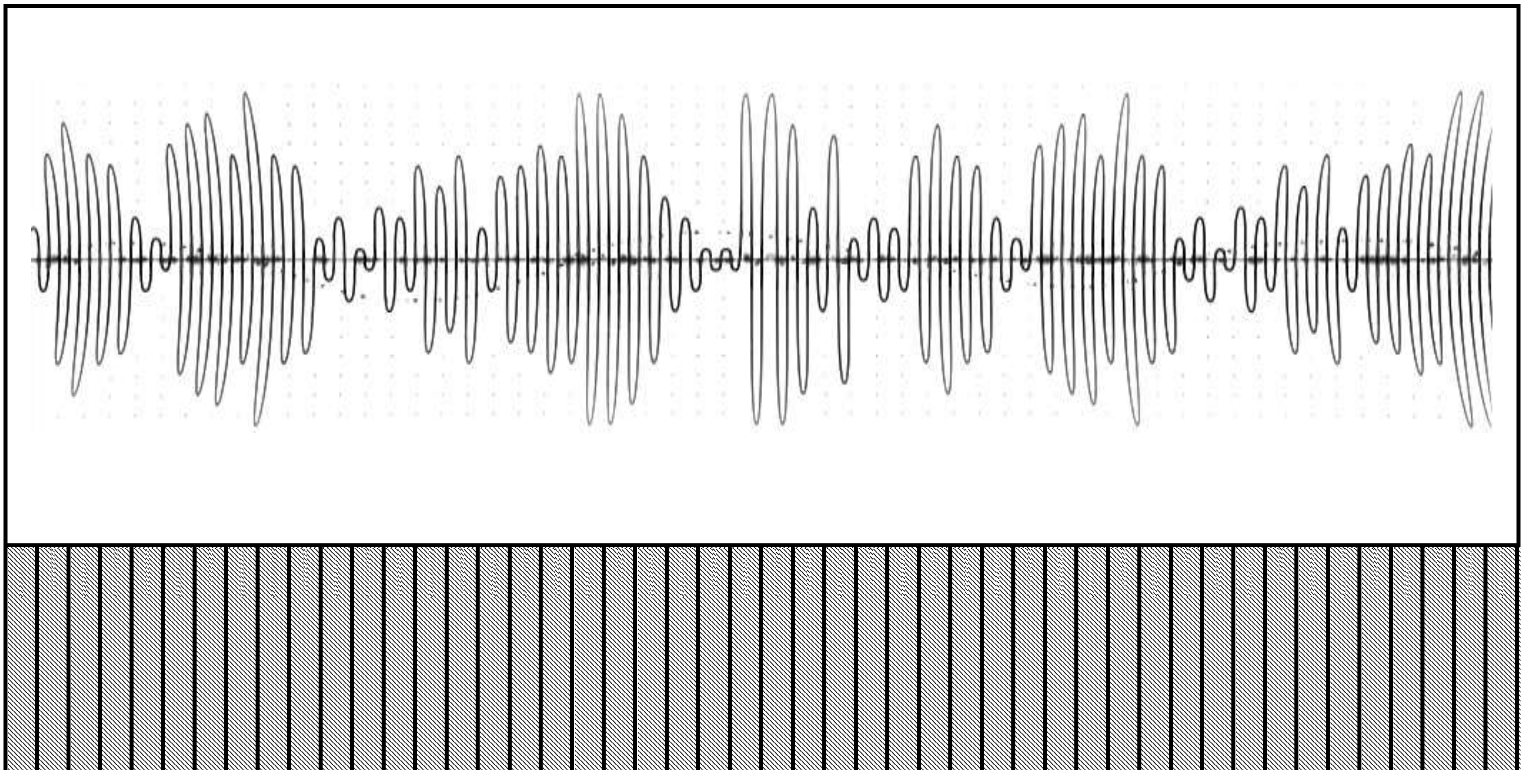
3.2.4 Stage Four: Merging

Merge the selected block frames in stage two with the sound data segments in stage three to construct a new summarized video.

- ❖ To clarify the processing operations that are doing in the proposed video summarization method, a simple example will be present below:

<u>Stage (1):</u>

- Partition the video frames and sound (**1…N**) into set of partitions **(1…M**)
- Each set of frames are divided based on five seconds **(it was taken every 5 seconds assumption)**.
- **M = N/5**.



5 sec	**5 sec**	**5 sec**	**5 sec**	**5 sec**	**5 sec**	**5 sec**	**5 sec**	**5 sec**	**5 sec**
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**

Figure (3-2): Video content

Stage (2): Video (Frames) Processing

Find Average Frame AFi for all the frames in each 5 seconds.

Table (3-1): Frames Processing

Frame	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Information of Frame	A B C D E F G H I J	A K C L E M G H I N	A K C B E M G H I D	P V C B E R S H I T	P W C U K X S D Z Y	V W C U H X R D B Y	V M E U H X R L B Y	V M E U H F R L B Y	V A E U H F G C B D	I A E J H F G C B D

- These characters **(A,B,C,D,…..)** were taken it on the information contained in each frame.

1. Calculate the difference ratio **(DR)** between each successive average frames$\mathbf{DR_{i,i+1}}$

Table (3-2): Ratio Finding

Frame	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Information of Frame	A B C D E F G H I J	A K CL EM GH IN	A K C B E M G H I D	PV C B E R S H I T	PW C U KX SD ZY	V W C U H X R D B Y	V M E U H X R L B Y	V M E U H F R L B Y	V A E U H F GC B D	I A E J H F G C B D

40% 20% 50% 70% 40% 30% 10% 40% 20%

- It was measured the difference between the frames by finding changes in characters between frame and other.

  For example: in **frame 1** and **frame2** the founding ratio **(40%)**

Frame	**1**	**2**
Information of Frame	A B C D E F G H I J	A K CL EM GH IN

The symbol **A** was repeated in two frames.**(There are similarities between the adjacent frames)**

The symbol **C** was repeated in two frames. **(There are similarities between the adjacent frames).**

The symbol **E** was repeated in two frames. **(There are similarities between the adjacent frames)**.

The symbol **G** was repeated in two frames. **(There are similarities between the adjacent frames).**

The symbol **H** was repeated in two frames. **(There are similarities between the adjacent frames)**.

The symbol **I** was repeated in two frames. **(There are similarities between the adjacent frames)**.

The differences in two frames **(B,K) , (D,L) , (F,M) ,(J,N). (There is a difference in these characteristics in adjacent frame).**

- We found four differences in the symbols of ten so the ratio has become 40%.

2. Choose the most frequent ratio $\mathbf{MF_{Ratio}}$ =**MostFreq(**$\mathbf{DR_{i,i+1}}$) in (2)

   Selected ratio = 40 %

3. Choose the desired average frame $\mathbf{AF_{i+1}}$ when the ratio of difference between $\mathbf{AF_i}$ and $\mathbf{AF_{i+1}}$>= $\mathbf{MF_{Ratio}}$.

Table (3-3): Result of frames

Frame	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Information of Frame	A B C D EF G H I J	A K CL EM GH IN	A K C B E M G H I D	PV C B E R S H I T	PW C U K X SD ZY	V W C U H X R D B Y	V M E U H X R L B Y	V M E U H F R L B Y	V A E U H F GC B D	I A E J H F G C B D

- Frame (1) ↔ Frame (2) = 40% OK
- Frame (2) ↔ Frame (3) = 20% Cancel
- Frame (2) ↔ Frame (4) = 60% OK
- Frame (4) ↔ Frame (5) = 70% OK
- Frame (5) ↔ Frame (6) = 40% OK
- Frame (6) ↔ Frame (7) = 30% Cancel
- Frame (6) ↔ Frame (8) = 30% Cancel
- Frame (6) ↔ Frame (9) = 40% OK
- Frame (9) ↔ Frame (10) = 20% OK

4. Collect all the average frames that have been chosen in the table (3-4)

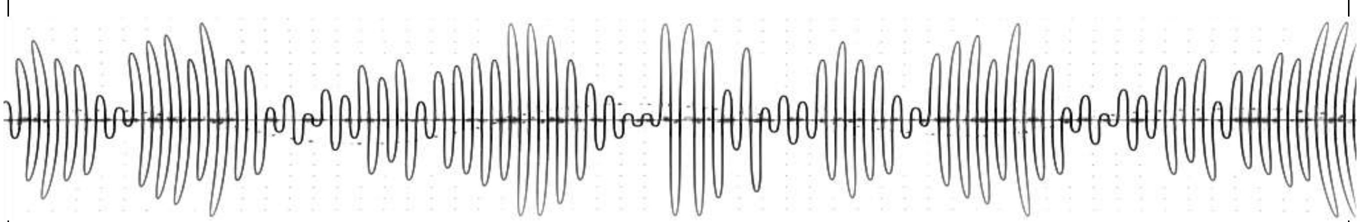
Table (3-4): Collecting of Average frames

Frame	**1**	**2**	**4**	**5**	**6**	**9**
Information of Frame	A B C D E F G H I J	A K C L E M G H I N	PV C B E R S H I T	PW C U K X SD ZY	V W C U H X R D B Y	V A E U H F GC B D

Stage (3): Sound Processing

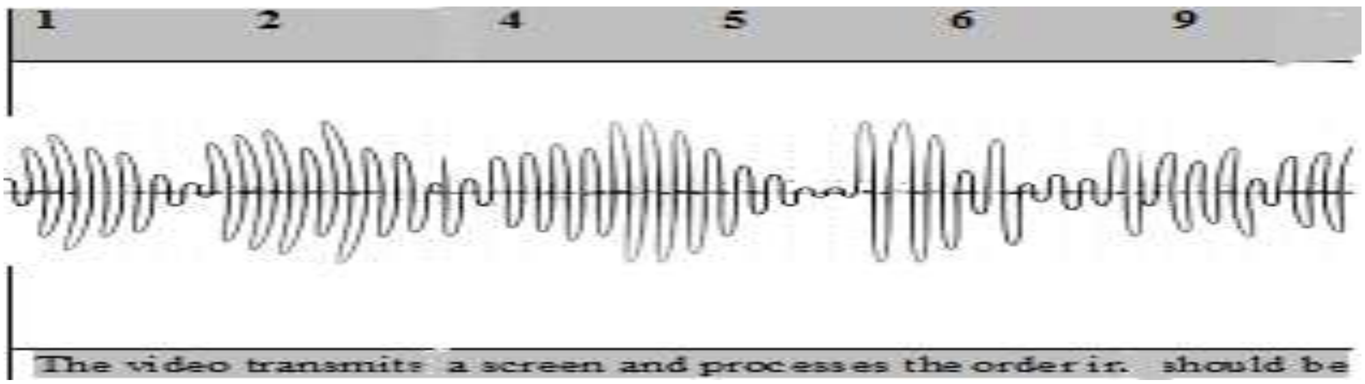
1. Find the suitable partition of sound data for each selected average frames in part (1)

Table (3-5): Sound Processing

Frame	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Signal Audio of Frame										
	The video transmits a signal to a screen and processes the order in which the screen captures should be shown.									

- We selected sound depending on the frame selected and was also taking in view of the existing words , there are some words are cut because they are not whole words .

For example, the word **captures** was cut, Knowing that the letters **es** within the frame 9 does not give meaning has been neglected.



<u>Stage (4):</u> Merge the selected frames from part (1) with the selected sound data from part (2) to construct new video file.

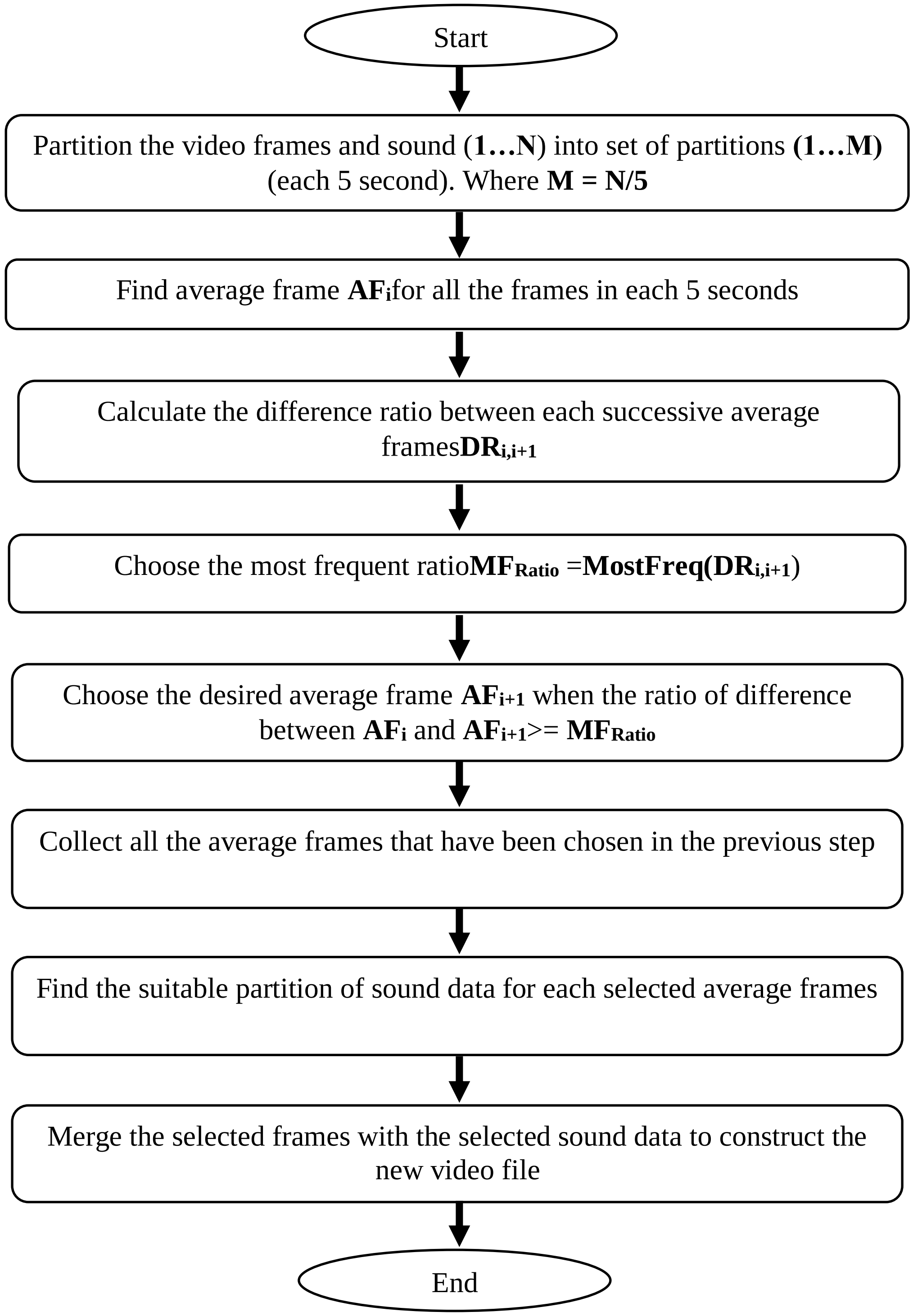


Figure (3-3): Flowchart of the proposed model

3.3Power Consumption and Bandwidth Model

In this work, power of the input video and summarized video can be computed using general equation of power (Proakis,2010)

$$P = \frac{1}{M*N} \sum_{Rows} \sum_{coulmns} x^2 \text{ ....(4)}$$

Where M and N are size of frame, x value of pixel after normalization. The general model of power can be shown in the following procedure

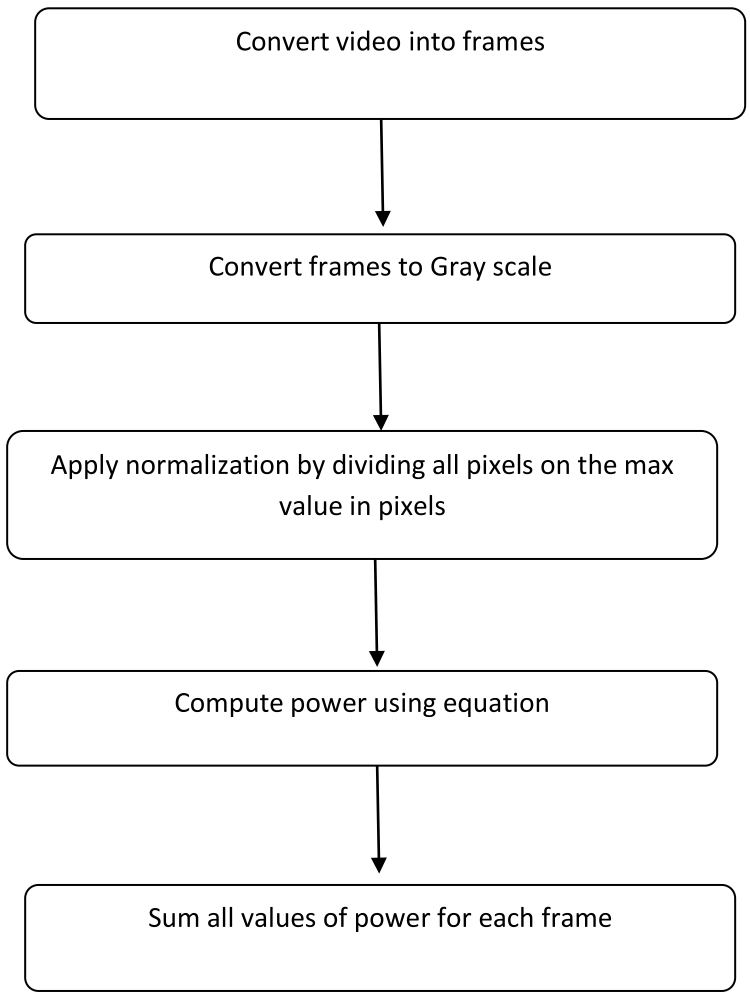


Figure (3-4): Flowchart of the power model

After computing total power, it can be used to compute required bandwidth that needed to transmit the video, assuming that total capacity 'C' of the channel is 10M bits/sec/Hz (Prakism,2010)

$$C = BW log_2(1 + P) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (5)$$

From the above equation, the required bandwidth can be computed as

$$BW = C/log_2(1 + P) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots .. (6)$$

From the above equation, it can be shown that once power decreased the required bandwidth increased, so the expected relation between power and bandwidth at certain capacity.

Chapter Four

Experiments And Results

4.1 Introduction

This chapter provides details of the performed experiments of the proposed methodology of testing 8 different videos; results obtained were then analyzed. The proposed methodology and framework were implemented using MATLAB. In each video, power before and after summarization is achieved and bandwidth before and after summarization is shown, also power spectral density (PSD) is illustrated for each video.

4.2 Data Set

The whole data set, used for experiments, is available to download through openvideo.org. This dataset makes recommendation for full understanding of the capabilities and specifics of the standard. The Open Video data set experimental results can be used as a test set, enabling different methods to be compared based in the same data set.

4.3 Experimental Setup

The experiments have been executed on a computer with the following technical features:

_ Operative System: Windows 7.

_ Processor: Intel Core I 3.

_ Memory: 4 GB.

_ Programming Language: MATLAB

4.4 Experiments

In this section, eight test videos will be illustrated. In each test videos, three figures will be shown. The first figure (4-1) represents the PSD of the original video presented in 3-D format, where x-axis and y-axis represent number of rows and columns of the frames in the video respectively. The z-axis represents the value of PSD, which is the equal total power dived by the area (number of rows times number of columns). In the figure, maximum of PSD is shown in red color.

The second figure (4-2) represents the power distribution of each tests, where x-axis is the total number of frames and y-axis is the power at each frame. While thirst figure repower the power after summarization where un-necessary frames were dropped and summarized, so the ripple for each parts of power is reduced as in the third figure (4-3).

It is evident that the numbers of frames are reduced before and after summarization as shown in x-axis for power in (4-2) and (4-3) figure.

Finally, each test video is different from other videos in terms of frame rate and bandwidth before summarization. Furthermore, duration for each video is different. The information contained in each video is also different, for example some videos contain only one person, while other videos contain more than one person.

Test1: this video is of 3.60 MB with 25 frames/second

In this case, PSD is shown in Figure (4-1), where PSD is derived using fft shift in MATLAB.

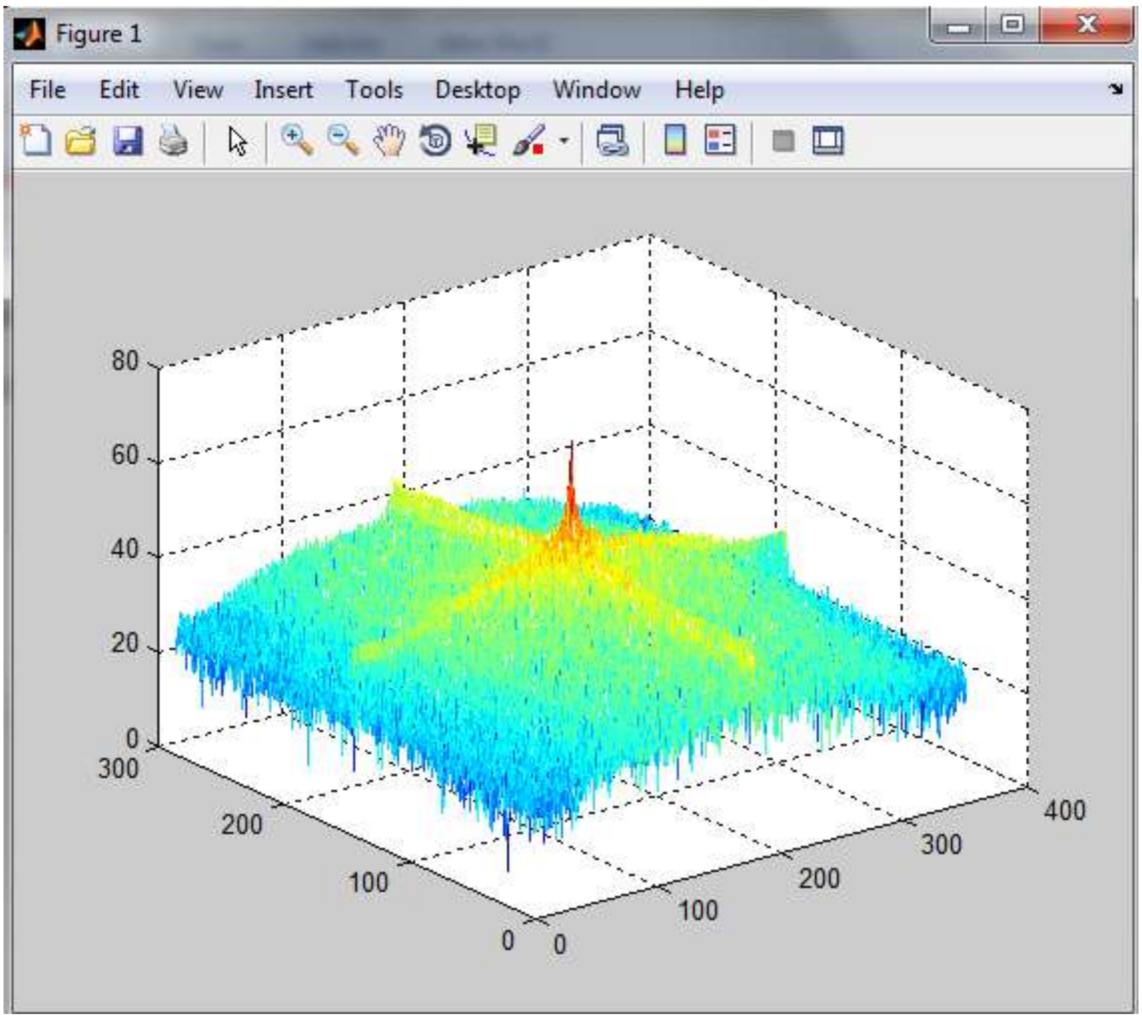


Figure (4-1): PSD for Test1

In above figure we have 300 frames at x-axis with average centralized power equal 30. Fig (4-2) and Fig. (4-3) show power data before and after summarization respectively. As shown below, the two figures follow each other, while in video summarization case, ripple is less compared to original video, this is excepted since video summarization drop certain frames which causes less ripple.

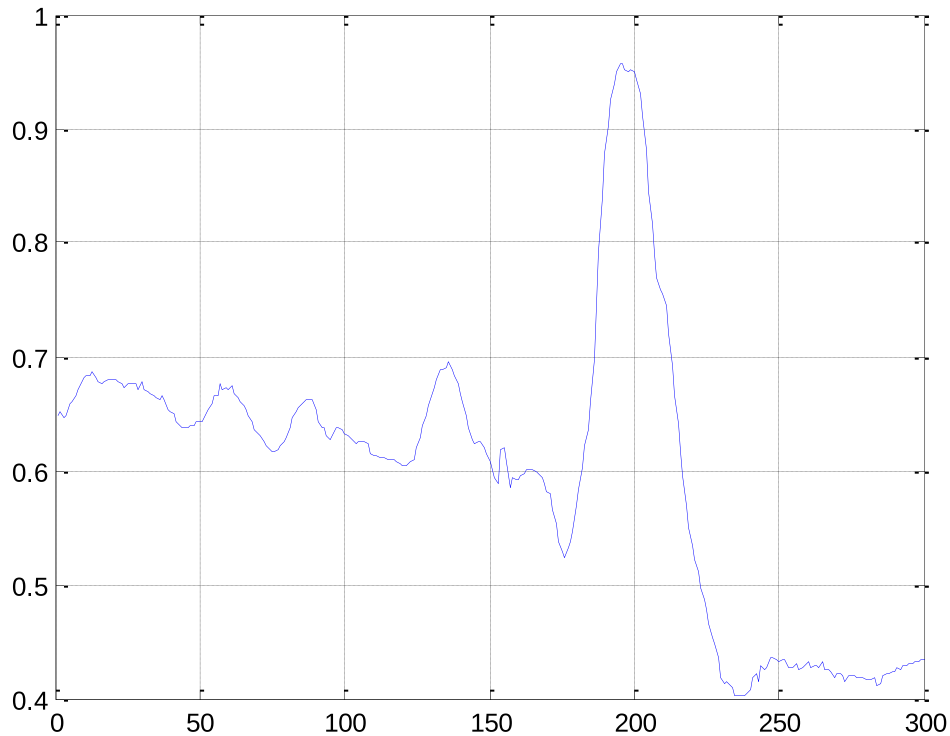


Figure (4-2): Power for Test1

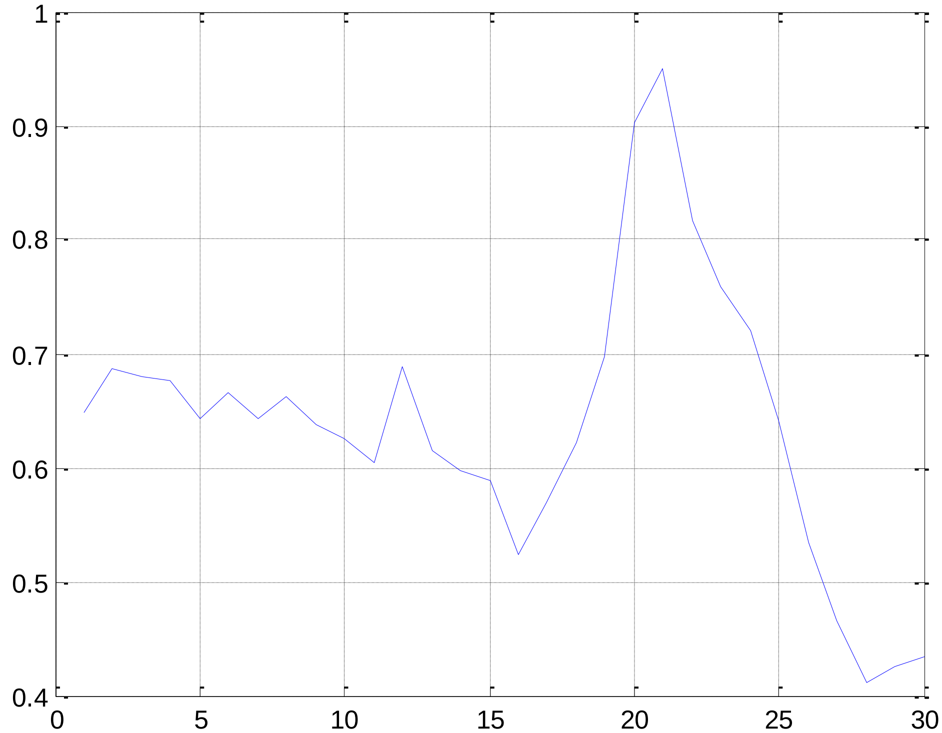


Figure (4-3): Power after summarization for Test1

Test2: this video is of 71.1MB with 30 frames/second.

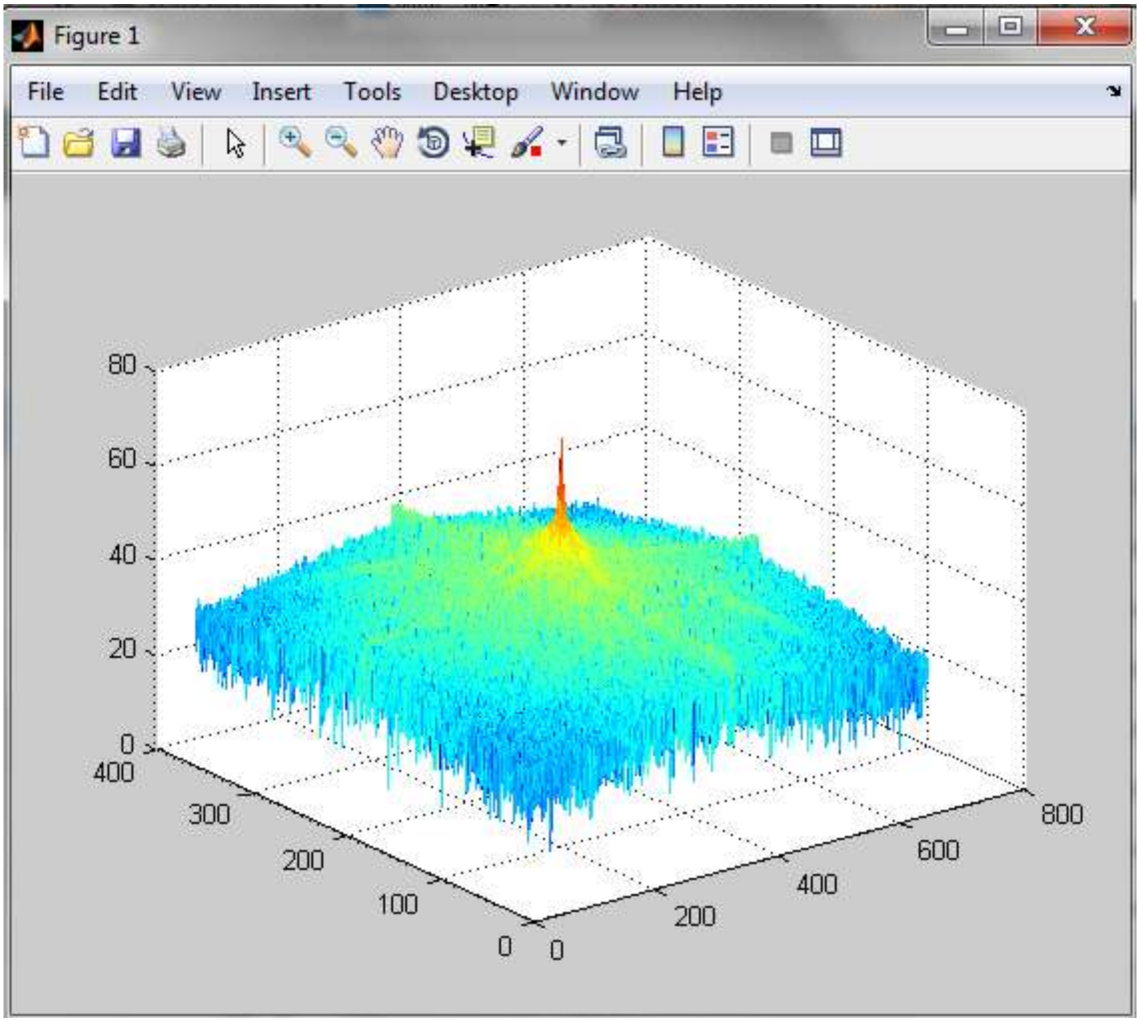


Figure (4-4): PSD for Test2

Fig (4-5) and Fig. (4-6) show power data before and after summarization respectively. As shown below, the two figures follow each other, while in video summarization case, ripple is less compared to original video, this is excepted since video summarization drop certain frames which causes less ripple

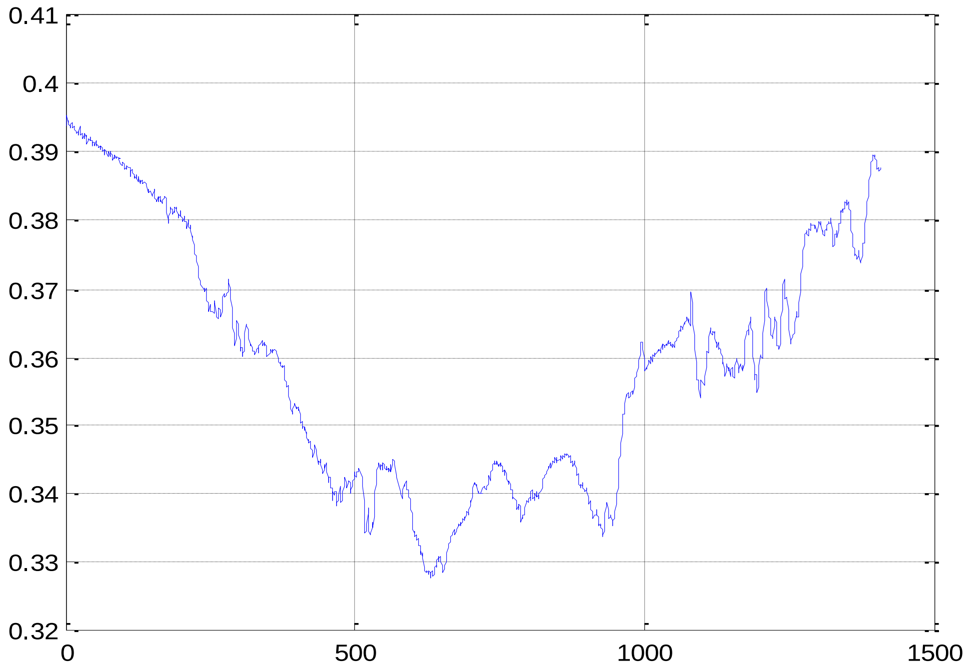


Figure (4-5): power for Test2

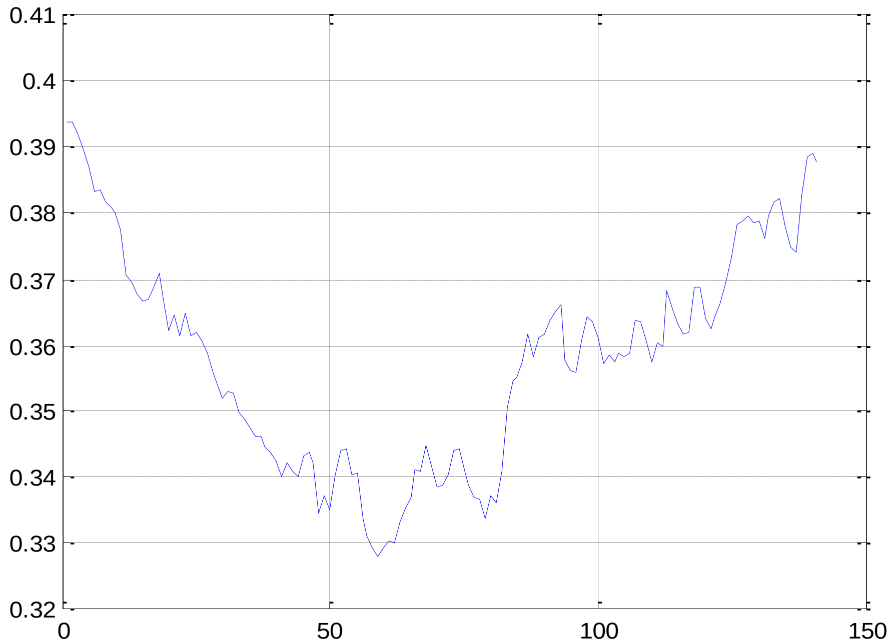


Figure (4-6): power after summarization for Test2

The following table illustrate power consumption for original video (P1) and its bandwidth (BW1) compared with power consumption of video summarization and bandwidth (P2,BW2).

Table (4-1): power consumption and bandwidth before and after summarization

Test #	P1	BW1	P2	BW2
Test 1	180.2067	1.33E+06	19.1337	2.31E+06
Test 2	506.2093	1.11E+06	50.4662	1.76E+06
Test 3	229.4493	1.27E+06	22.9293	2.18E+06
Test 4	86.153	1.55E+06	8.4746	3.08E+06
Test 5	60.0513	1.69E+06	6.2879	3.49E+06
Test 6	160.6731	1.36E+06	15.7844	2.46E+06
Test 7	134.9681	1.41E+06	14.3174	2.54E+06
Test 8	200.7706	1.31E+06	19.4922	2.30E+06

In order to compare between effect on the bandwidth before and after summarization, the following chart illustrate, the increasing in bandwidth for each test video.

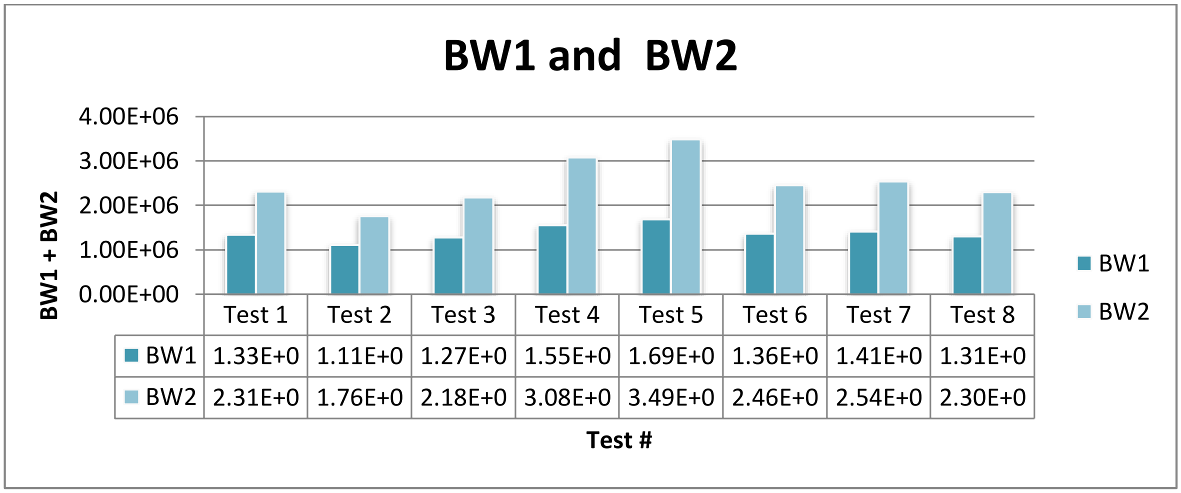


Figure (4-7): BW1 and BW2

In order to compare between effect on the power before and after summarization, the following chart illustrate, the decreasing in power for each test video.

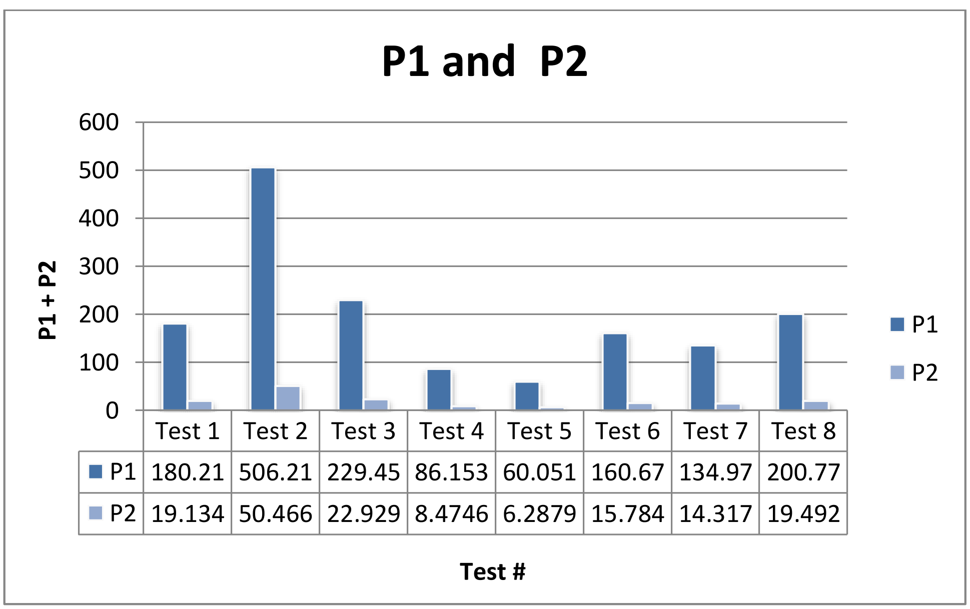


Figure (4-8): P1 and P2

Last observation about the results is to compare how much the reduction in time before and after the video data, In order to compare between effect on the time before and after summarization, the following chart table, the deceasing in time for each test video.

Table (4-2): Time before and after summarization

Test #	Duration (sec)
Test 1	**12**
Test 1_summary	**2**
Test 2	**47**
Test 2_summary	**9**
Test 3	**13**
Test 3_summary	**2**
Test 4	**4**
Test 4_summary	**1**
Test 5	**5**
Test 5_summary	**1**
Test 6	**11**
Test 6_summary	**2**
Test 7	**9**
Test 7_summary	**2**
Test 8	**20**
Test 8_summary	**4**

Finally, we can plot the relation between decreasing power and increasing in bandwidth for first test video, assuming bandwidth is normalized to bw1=1.33e2 and bw2=2.31e2;

The following figure, the x-axis means just the index to plot the two figures, while y-axis represent power and bandwidth values.

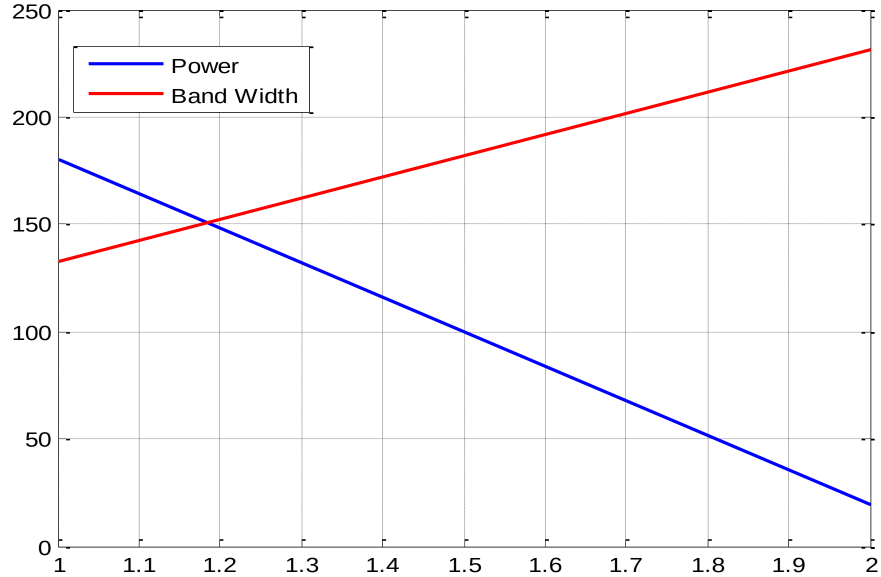


Figure (4-9): Power and bandwidth relation

In case of video with audio data, the system will find the suitable partition of sound data for each selected average frames. The following figures show the audio data as function of time.

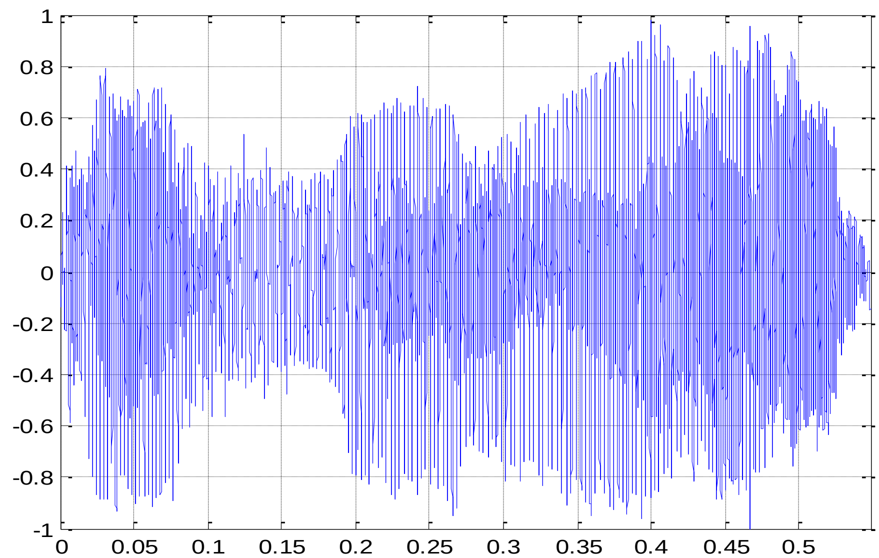


Figure (4-10): audio signal after filtration and removing partitions

The spectrum of the audio can be shown below, where it represent the time domain and frequency domain of the audio, where x-axis is the time and y-axis is the frequency, as shown main data is concentrated in range zero to 0.5 Hz.

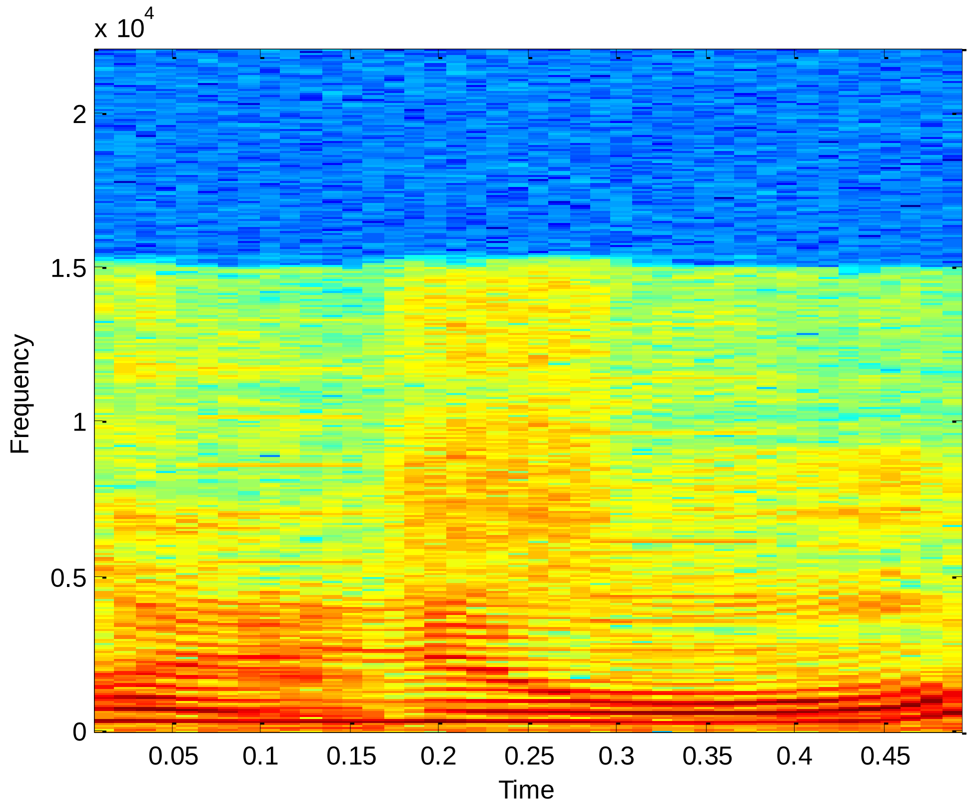


Figure (4-11): spectrogram of the audio data

The main frames of the summarized video for one of the test videos, is shown below

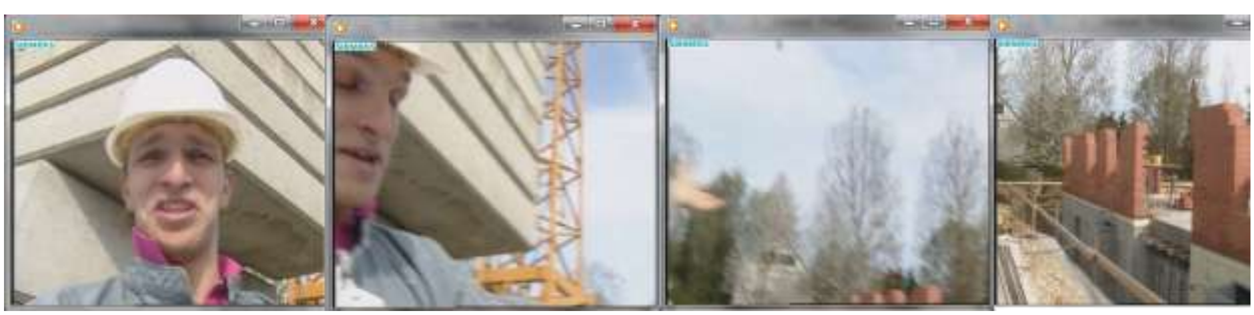


Figure (4-12): Summarized video frames

4.5 Evaluation

According to (Jadhav and Jadhav, 2015) , the authors proposed Video Summarization using High Order Colour Moments (VSUHCM), where VSUHCM method proposed in the paper can be divided into two: the Key Frame Extraction and the Shot Boundary Detection. Shot boundary detection method summarizes video in segments of multiple shots. In this method, the image histogram functions like the graphical representation of digital image tonal distribution. In this method, there are eight steps, they include frame extraction from video-frame partition into m*n block, computation of Kurtosis (K), Skewness (S), Standard deviation (S), and Mean, by use of image histogram. The fourth step is computing the difference, then adding all the four differences followed by calculation of Td using step 4 and 5 for the consecutive frames. Step seven involves calculation of the threshold and then comparing the threshold T with TD (Total Difference) and finally, if the Td $(i+1) \geq T$, then it marks the end of the shot before and start the following shot (i+1).

The key frame extraction, on the other hand, has three major steps which include: firstly the entire video sequence division, then fining the frame that consists of the highest mean from each shot to form static summarization which is the last step. Same test video is applied according to the proposed method in this thesis, where PSD and power before and after are shown in the figures below.

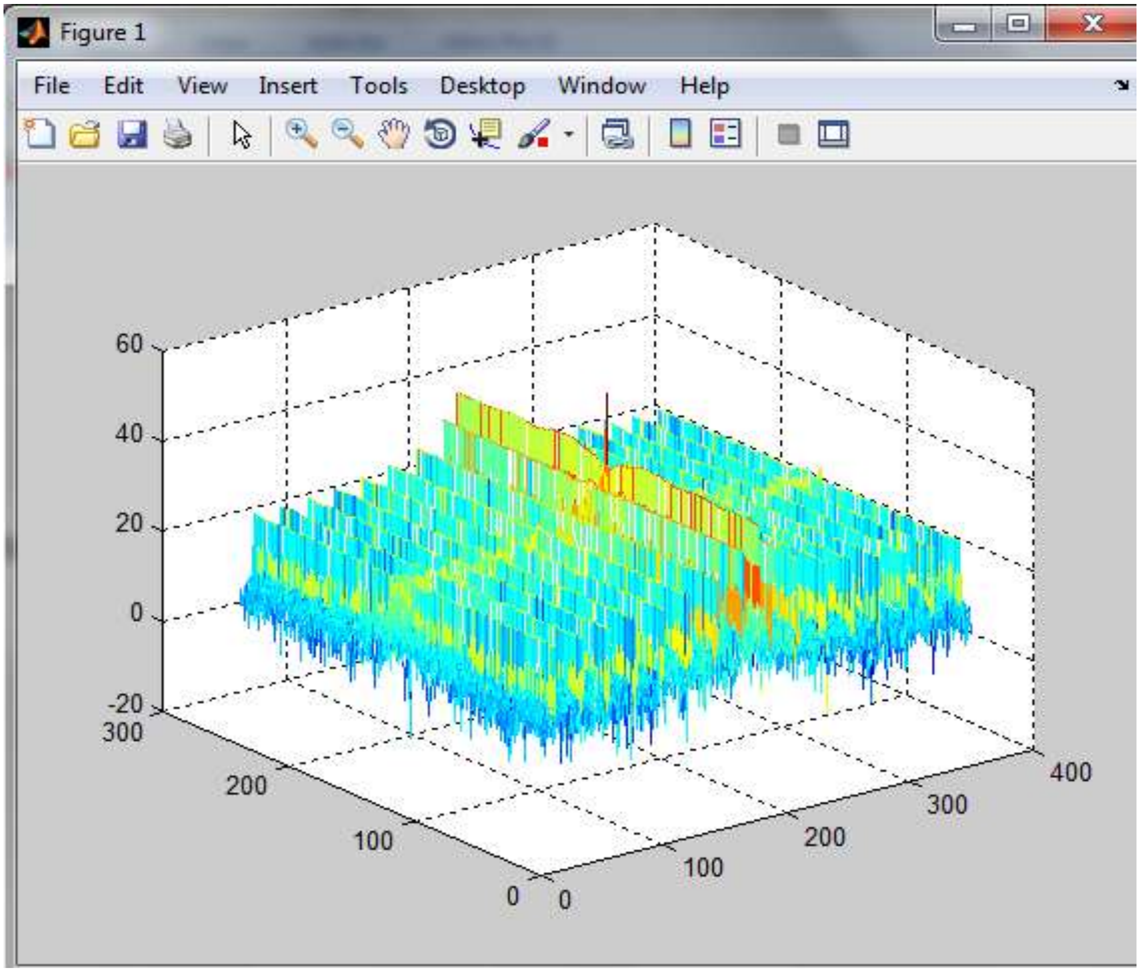


Figure (4- 13): PSD for Jadhav and Jadhav

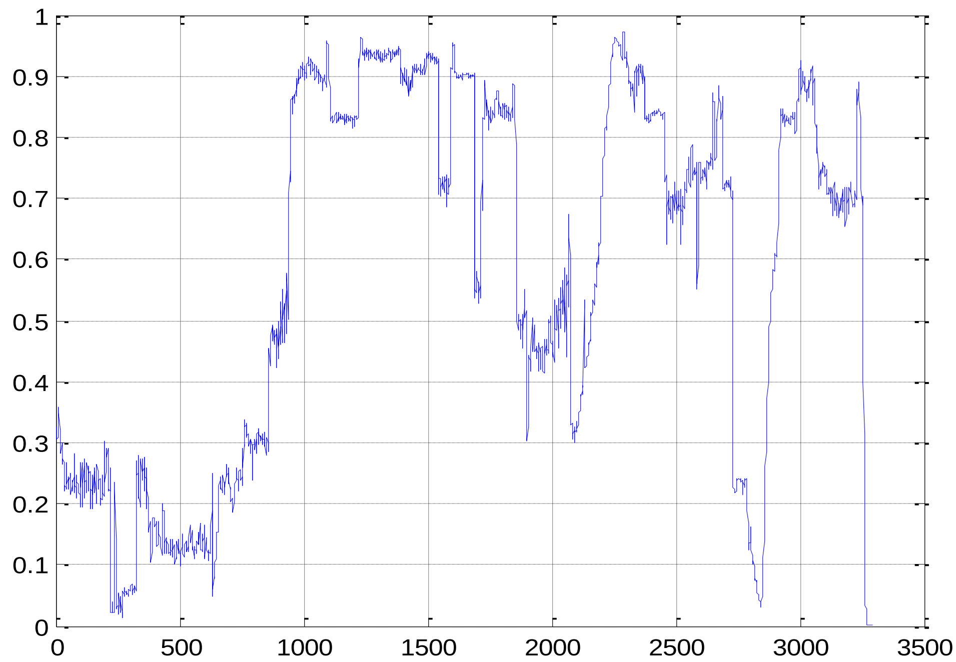


Figure (4-14): Power for Jadhav, &Jadhav

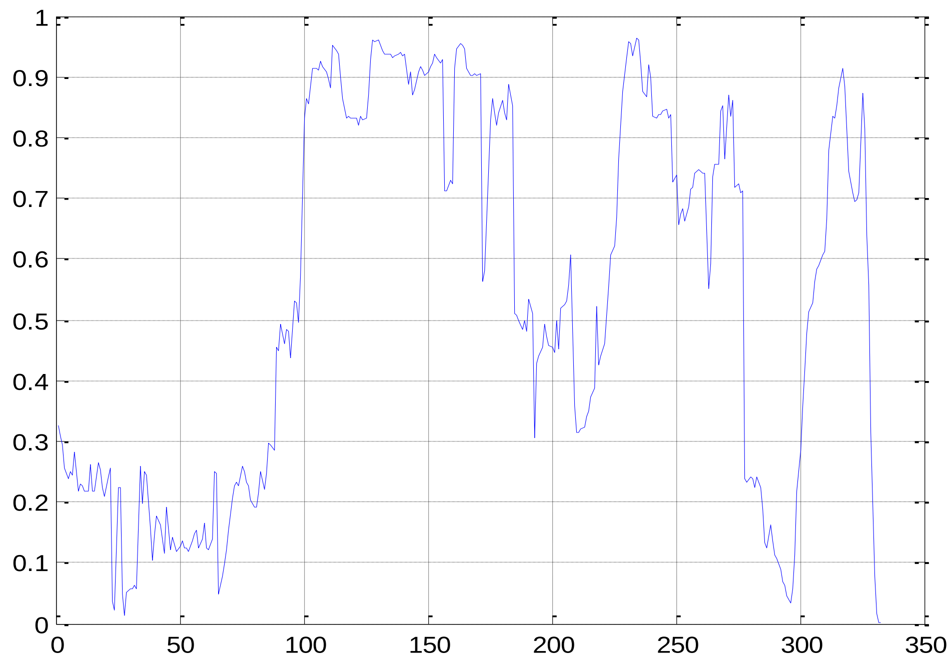


Figure (4- 15): Power After Summarization for Jadhav and Jadhav

When compared between the research of Jadhav and Jadhav and the methodology proposed in this thesis, different points can be achieved:

- In Jadhav and Jadhav work, output video or frames are smooth while in proposed method output video contains some missed frames that not provide overall information about video.
- In Jadhav and Jadhav work, frames of output video are taken within specific periods according to the algorithm which increase the processing time and complexity, while in the proposed work the complexity of the proposed algorithm is simple compared to other methods.
- In Jadhav and Jadhav work, dynamic video summarization is used while in the proposed method static video summarization is used.

According to (Jadhav and Jadhav, 2015), from the experiments the results from the quantitative analysis and qualitative analysis indicated that the method gave the best result with high video quality. In addition, the qualitative analysis method used indicated that the algorithm helped in capturing a distribution of color that is more detailed from the image. The qualitative analysis showed videos in high definition. To enhance the achievement of video skimming, the proposed method's future scope will test on various genre videos such as TV-shows, sports, cartoons and talk shows. The figure below, illustrate frame after summarization according to (Jadhav and Jadhav, 2015), for different methods.

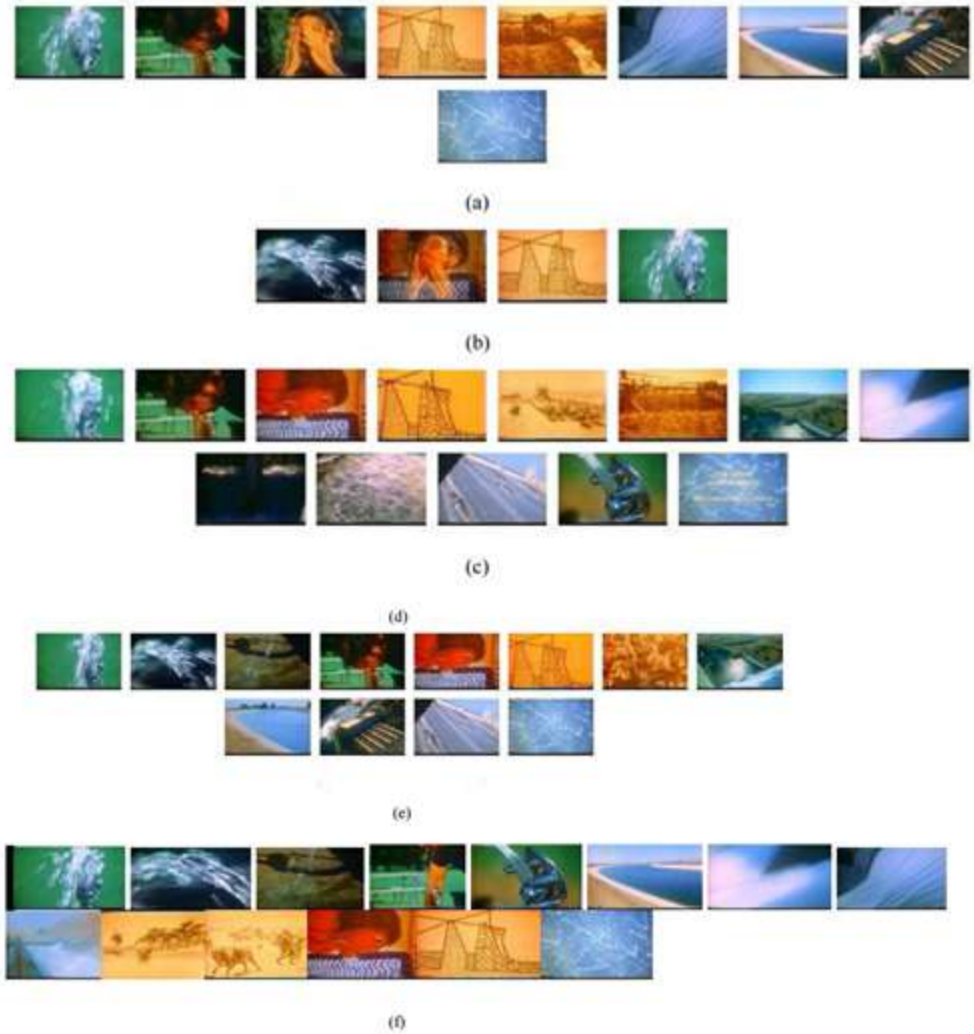


Figure (4-16) Video Summary result of

(a) OV (b) DT (c) VISTO (d) VSUMM1 (e) VSUMM2 (f) VSUHCM

In order to compare with (Jadhav and Jadhav, 2015), same video is applied to study the performance of video after summarization, as shown below the snapshot of video after summarization, illustrate that the proposed method can be used for different points in the video especially when moving from view to different view, the proposed method, also illustrate that there is no any redundancy in the frames, which provide high efficiency in the ratio between video before and after summarization.

Figure (4-17): Video Summary Result Using Proposed Method

Chapter FIVE

Conclusion and Future Work

5.1 Conclusion

Video summarization has recently attracted significant research. Accordingly mechanisms and techniques have been suggested. In this thesis, we presented the effectiveness of our investigation through focusing on reducing power consumption in mobile applications and additional cost in the process of downloading a certain video signal. Therefore, a video summarization is being applied on 8 test videos, a remarkable result regarding both power and bandwidth is achieved as the power consumption is been reduced by 80% and the bandwidth is being increased by 40%.

Moreover, the results show that there is a tradeoff between power consumption and bandwidth. Since, when reducing power according to Shannon theorem, the required bandwidth is increased in order to get constant capacity of the overall system.

In this thesis we first split the original video into video and audio contents using proposed algorithm mentioned in chapter three, the video is being divided into its contained frames and each frame is divided according to a time factor (here we take 5 seconds assumption), then we take an average for the set of block frames.

Regarding the audio part, we find the suitable partition for the selected average frame mentioned previously; although in some cases we disabled some words or voice segments if it is not suited (a word does not appear as a whole word).Finally, we merge the output from the previous steps to construct the new processed video in some cases an unnecessary audio signal is being found, we remove it by searching on zero crossing of audio signal and delete that range.

5.2 Future Work

Different research directions can be added to this research as a future work

- Studying effect of size of the video to apply efficient video summarization by applying optimization method to work for different video size even for long size video.
- Studying the effect of quality of the video, by summarization bad quality video.

Reference

Almeida, J., Leite, N. J., & Torres, R. D. S. (2012). Vison: Video summarization for online applications. *Pattern Recognition Letters*, *33*(4).

Almeida, J., Leite, N. J., & Torres, R. D. S. (2013).Online video summarization on compressed domain. *Journal of Visual Communication and Image Representation*, *24*(6).

Cahuina, E. J. C., & Chavez, G. C. (2013, August). A new method for static video summarization using local descriptors and video temporal segmentation. In *2013 XXVI Conference on Graphics, Patterns and Images* (pp. 226-233). IEEE.

Dang, C. T., & Radha, H. (2014). Heterogeneity image patch index and its application to consumer video summarization. *IEEE Transactions on Image Processing*, *23*(6), (pp.2704-2718).

Elkhattabi, Z., Tabii, Y., & Benkaddour, A. (2015). Video Summarization: Techniques and Applications. *World Academy of Science, Engineering and Technology, International Journal of Computer, Electrical, Automation, Control and Information Engineering*, *9*(4), (pp. 891-896).

Furini, M., Geraci, F., Montangero, M., Pellegrini, M., 2010. STIMO: STIll andMOvingvideo storyboard for the web scenario. *Multimedia Tools Appl*

Gašić, M., Breslin, C., Henderson, M., Kim, D., Szummer, M., Thomson, B., & Young, S. (2013). On-line policy optimisation of Bayesian spoken dialogue systems via human interaction.In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 8367-8371).

Jadhav, M. P. S., & Jadhav, D. S. (2015).Video Summarization Using Higher Order Colour Moments (VSUHCM). *Procedia Computer Science*, *45*, (pp. 275-281).

Khan,S.,Pawar,Soudamini. (2015). Video Summarization: Survey on Event Detection and Summarization in Soccer Videos. *International Journal of Advanced Computer Science and Applications.*

Khosla, A., Hamid, R., Lin, C. J., & Sundaresan, N. (2013).Large-scale video summarization using web-image priors.*In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2698-2705).

Knox, T. B., Juma, E. O., Ochomo, E. O., Jamet, H. P., Ndungo, L., Chege, P., & Coetzee, M. (2014). An online tool for mapping insecticide resistance in major Anopheles vectors of human malaria parasites and review of resistance status for the Afrotropical region. *Parasites & vectors*,*7*(1).

Kokaram, A. C. (2013). *Motion picture restoration: digital algorithms for artefact suppression in degraded motion picture film and video*. Springer Science & Business Media.

Kumar, K., Liu, J., Lu, Y. H., & Bhargava, B. (2013).A survey of computation offloading for

Lee, Y. J., Ghosh, J., & Grauman, K. (2012). Discovering important people and objects for egocentric video summarization. *In CVPR* (Vol. 2, No. 6, p. 7).

Liu, g., & Zhao, j. (2010). Key frame extraction from mpeg video stream. In information processing (isip). *Third international symposium on Ieee* (pp. 423-427).

Lu, Z., & Grauman, K. (2013).Story-driven summarization for egocentric video.In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (pp. 2714-2721).

Mayberry, T., Blass, E. O., & Chan, A. H. (2014).Efficient Private File Retrieval by Combining ORAM and PIR. In *NDSS*.

mobile systems. *Mobile Networks and Applications*, 18(1), (pp. 129-140).

Mundur, P., Rao, Y., Yesha, Y., 2006. Keyframe-based video summarization usingDelaunay clustering. *Internat. J. Dig. Libr*. 6 (2), (pp. 219-232).

Potapov, D., Douze, M., Harchaoui, Z., & Schmid, C. (2014).Category-specific video summarization.In *European conference on computer vision* (pp. 540-555).Springer International Publishing.

Rehman, A., & Saba, T. (2014).Features extraction for soccer video semantic analysis: current achievements and remaining issues. *Artificial Intelligence Review*, *41*(3), (pp. 451-461).

Saba, T., & Altameem, A. (2013). Analysis of vision based systems to detect real time goal events in soccer videos. *Applied Artificial Intelligence*, *27*(7), (pp. 656-667).

Schuster, G. M., & Katsaggelos, A. (2013). Rate-Distortion based video compression: optimal video frame compression and object boundary encoding. *Springer Science & Business Media.*

Sigari, M. H., Soltanian-Zadeh, H., & Pourreza, H. R. (2015). Fast highlight detection and scoring for broadcast soccer video summarization using on-demand feature extraction and fuzzy inference. *International Journal of Computer Graphics*, *6*(1).

Zhang, L., Xia, Y., Mao, K., Ma, H., & Shan, Z. (2015).An effective video summarization framework toward handheld devices. *IEEE Transactions on Industrial Electronics*, *62*(2), (pp. 1309-1316)