AUTOMATIC VIDEO CATEGORIZATION AND SUMMARIZATION

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

KEZBAN DEMİRTAŞ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
COMPUTER ENGINEERING

SEPTEMBER 2009

Approval of the thesis:

AUTOMATIC VIDEO CATEGORIZATION AND SUMMARIZATION

submitted by **KEZBAN DEMİRTAŞ** in partial fulfillment of the requirements for the degree of **Master of Science in Computer Engineering Department, Middle East Technical University** by,

Prof. Dr. Canan Özgen _____
Dean, Graduate School of **Natural and Applied Sciences**

Prof. Dr. Müslim Bozyiğit _____
Head of Department, **Computer Engineering**

Assoc. Prof. Dr. Nihan Kesim Çiçekli _____
Supervisor, **Computer Engineering Dept., METU**

Assoc. Prof. Dr. İlyas Çiçekli _____
Co-Supervisor, **Computer Engineering Dept., Bilkent University**

Examining Committee Members:

Assoc. Prof. Dr. Cem Bozşahin _____
Computer Engineering Dept., METU

Assoc. Prof. Dr. Nihan Kesim Çiçekli _____
Computer Engineering Dept., METU

Assoc. Prof. Dr. Ferda Nur Alparslan _____
Computer Engineering Dept., METU

Dr. Ayşenur Birtürk _____
Computer Engineering Dept., METU

Dr. Orkunt Sabuncu _____
Orbim, TEKNOKENT

Date: 10 / 09 / 2009

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name : Kezban Demirtaş

Signature :

ABSTRACT

AUTOMATIC VIDEO CATEGORIZATION AND SUMMARIZATION

Demirtaş, Kezban

M.S., Department of Computer Engineering

Supervisor: Assoc. Prof. Dr. Nihan Kesim Çiçekli

Co-Supervisor: Assoc. Prof. Dr. İlyas Çiçekli

September 2009, 86 pages

In this thesis, we make automatic video categorization and summarization by using subtitles of videos. We propose two methods for video categorization. The first method makes unsupervised categorization by applying natural language processing techniques on video subtitles and uses the WordNet lexical database and WordNet domains. The method starts with text preprocessing. Then a keyword extraction algorithm and a word sense disambiguation method are applied. The WordNet domains that correspond to the correct senses of keywords are extracted. Video is assigned a category label based on the extracted domains. The second method has the same steps for extracting WordNet domains of video but makes categorization by using a learning module. Experiments with documentary videos give promising results in discovering the correct categories of videos.

Video summarization algorithms present condensed versions of a full length video by identifying the most significant parts of the video. We propose a video summarization method using the subtitles of videos and text summarization techniques. We identify significant sentences in the subtitles of a video by using text summarization techniques and then we compose a video summary by finding the video parts corresponding to these summary sentences.

Keywords: Video Categorization, Video Summarization, Text Summarization, WordNet Domains

ÖZ

VİDEOLARIN OTOMATİK OLARAK SINIFLANDIRILMASI VE ÖZETLENMESİ

Demirtaş, Kezban

Yüksek Lisans, Bilgisayar Mühendisliği Bölümü

Tez Yöneticisi: Doç. Dr. Nihan Kesim Çiçekli

Ortak Tez Yöneticisi: Doç. Dr. İlyas Çiçekli

Eylül 2009, 86 sayfa

Bu tezde, videoların, altyazılarını kullanarak otomatik sınıflandırılması ve özetlenmesi gerçekleştirilmektedir. Videoların sınıflandırılması için iki metod önermekteyiz. İlk metod, video altyazılarına doğal dil işleme tekniklerini uygulayarak ve WordNet sözlük veritabanını ve WordNet alanlarını kullanarak otomatik kategorileme yapmaktadır. Metod yazı ön işlemesi ile başlar. Daha sonra anahtar sözcük çıkarma ve kelime anlamını ayırt etme metodları uygulanır. Anahtar kelimelerin ayırtedilen anlamlarına denk gelen WordNet alanları çıkarılır. Çıkarılan bu WordNet alanları temel alınarak videoya bir kategori etiketi atanır. İkinci metod da WordNet alanlarını çıkarmak için ilk metodla aynı adımları izler fakat bir öğrenme modülü kullanarak kategorileme gerçekleştirir. Belgesel videolarıyla yapılan

deneyler, videoların gerçek kategorilerinin bulunmasında başarılı sonuçlar vermiştir.

Video özetleme algoritmaları, videonun en önemli kısımlarını belirleyerek videonun özetlenmiş biçimlerini sunmaktadır. Biz de videoların altyazılarını ve yazı özetleme tekniklerini kullanan bir video özetleme metodu öneriyoruz. Yazı özetleme tekniklerini kullanarak, video altyazısının önemli cümlelerini belirliyoruz ve bu cümlelere denk gelen video kesitlerini bularak video özeti oluşturuyoruz.

Anahtar Kelimeler: Videoların Sınıflandırılması, Videoların Özetlenmesi, Yazı Özetlenmesi, WordNet Domainleri

To My Dear Family,

ACKNOWLEDGMENTS

I would like to express my deepest gratitude and profound respect to both my supervisor Assoc. Prof. Dr. Nihan Kesim Çiçekli and my co-supervisor Assoc. Prof. Dr. İlyas Çiçekli for their expert guidance and suggestions, positive approach throughout my masters study and their efforts and patience during supervision of the thesis.

I would like to express my sincere appreciation to the jury members, Assoc. Prof. Dr. Cem Bozşahin, Assoc. Prof. Dr. Ferda Nur Alparslan, Dr. Ayşenur Birtürk and Dr. Orkunt Sabuncu for reviewing and evaluating my thesis.

I would like to thank to TÜBİTAK UEKAE / İLTAREN for supporting my academic studies.

I would like to express my thanks to Mustafa Başıbüyük, for his endless patience, encouragement and support during my thesis work.

Finally special thanks to my family for bringing me up and making me who I am with their love, trust, understanding and every kind of support throughout my life.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

CHAPTER 1

INTRODUCTION

1.1 Video Categorization

Today people have access to a large amount of video from both television and internet. So finding a video of interest becomes a difficult and time-consuming job. It can be infeasible for a human to go through all available videos to find the video of interest. One method that people use to narrow down their choices is looking for video within specific categories. Because of large amount of video to categorize, research about automatic video categorization has begun [3, 9, 22, 32].

Video categorization algorithms assign video a meaningful label (e.g., "sports video" or "comedy video"). Most of the algorithms concentrate on classifying an entire video but some algorithms attempt to perform classification at the shot or scene level. Classifying at the shot or scene level can be useful for content filtering such as identifying violent or scary scenes in a movie or finding the news segments of an entire news broadcast. By this way, categories of videos can be subdivided, such as a category of action movies that include violent scenes.

In video categorization, generally video is classified into one of the several broad categories such as documentary type (e.g., geography, animals, religion) or movie genre (e.g., action, comedy, drama, horror) but sometimes

narrower categories can be defined such as specific types of sports video (e.g., tennis, basketball, football). The most popular domain for classification is entertainment video such as movies or sports but there exists some research about classifying informational video such as news or documentary.

For performing automatic classification of video, features are drawn from three modalities: visual, audio or text. Also some combination of these features can be exploited together. So video classification approaches could be divided into four groups: text-based approaches, audio-based approaches, visual-based approaches and those that use some combination of visual, audio and text features. Most of the approaches in the literature perform classification by utilizing features from a single modality because it is difficult to combine the features from three different areas.

Text-based approaches [2, 3, 4, 8, 38] are the least common in the video classification literature but have several benefits over other approaches. First of all, text processing is a more lightweight process than video and audio processing. Also text categorization techniques have been studied extensively in the computational linguistics literature. This accumulation can be exploited in video classification domain. Beside this, semantic of the video content is closely related to human language.

In this thesis, we propose two methods for video categorization. The first method, "Category Label Assignment", makes categorization by applying natural language processing techniques on video subtitles and uses the WordNet lexical database and WordNet domains. The method starts with text preprocessing. Then a keyword extraction algorithm and a word sense disambiguation method are applied. The WordNet domains that correspond to the correct senses of keywords are extracted. Video is assigned a category

label based on the extracted domains. The second method, "Categorization by Learning", has the same steps for extracting WordNet domains of video but makes categorization by using a learning mechanism. The overview of our video categorization system is given in Figure 1.

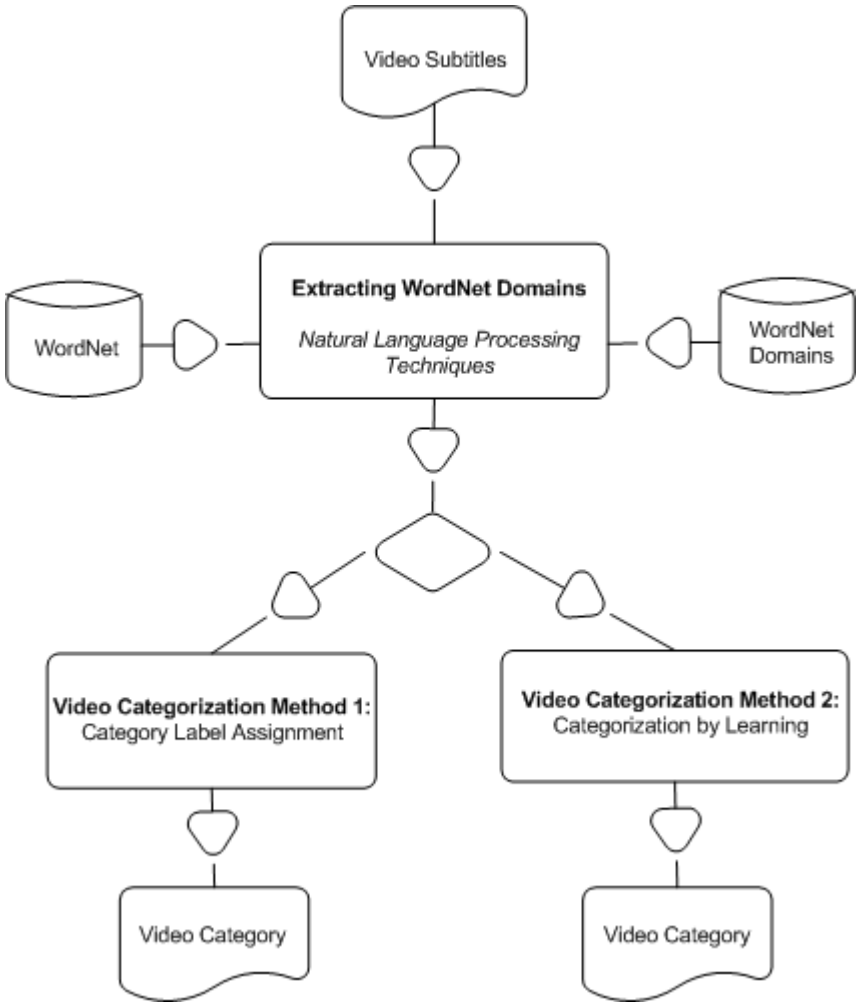

Figure 1 Video Categorization System

We select documentary videos as the classification domain and we predefine documentary categories as Geography, History, Animals, Politics, Religion, Sports, Music, Accidents, Art, Science, Transportation, Technology, People

and War. Experiments with both methods give promising results in discovering the correct categories of videos.

1.2 Video Summarization

With the increasing availability of digital video from both internet and television, users require assistance in accessing digital video. Video summarization algorithms present condensed versions of a full length video by identifying the most significant parts of the video. In order to have an idea about the content of a video, using such a summary is much easier than going through all the video.

In literature, there are two main trends in video summarization: still-image summaries and moving-image summaries. *Still-image summaries* are based on extracting individual key frames representing the content of the video in a static way [88, 89]. Generally video is segmented into shots and key frames representing these shots are selected to be included in the summary. *Moving-image summaries* are a collection of original video parts [90, 91]. These summaries can be classified into two sub-types: video previews and video summaries. Video previews present the most interesting parts of a video like a movie trailer, whereas video summaries keep the semantic meaning of the original video.

Video summaries are either used individually or integrated into various applications, such as browsing and searching systems. Such an application facilitates users managing and effectively accessing digital video content.

Video summarization systems produce summaries by analyzing the content of a video and condensing this content into an abbreviated form. Since video has a multimodal nature, video summarization can be performed by using

the image features, audio features or text features of video. Also some combination of these features can be exploited together.

In this thesis, we propose a video summarization system by using the text features of video and text summarization techniques. Text summarization techniques identify the significant parts of a text to constitute a summary. We select documentary videos as the summarization domain and we make use of documentary subtitles. We extract a summary of video text and then we find the video parts corresponding to these summary parts. By combining the video parts, we create a moving-image summary of the original video. The overview of our video summarization system is given in Figure 2.

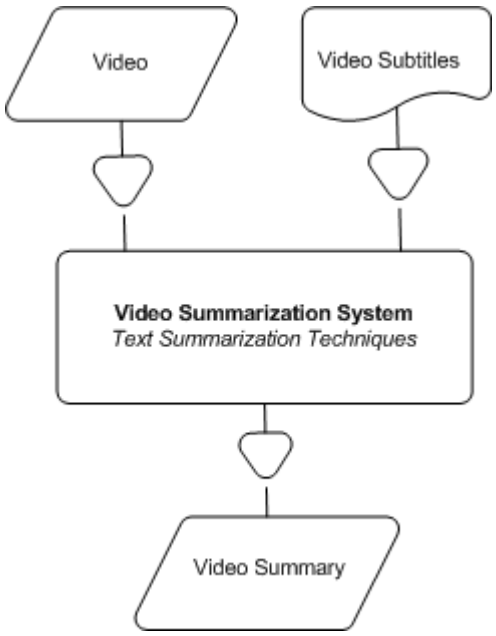

Figure 2 Video Summarization System

In our summarization approach, we take the advantage of the documentary videos characteristics. In documentary videos, speech is mostly composed of monolog and this monolog describes the things seen on the screen. For

example; in a documentary about “animals”, when an animal is seen on the screen, the speaker usually mentions that animal. So, when we find the video parts corresponding to the summary sentences of a video, those video parts are closely related with the summary sentences. Hence we obtain a semantic video summary giving the important parts of a video.

1.3 Thesis Goals

This thesis addresses two major problems:

1. Video Categorization

2. Video Summarization

For video categorization we propose two methods. The first categorization method is based on an existing video categorization algorithm (Katsiouli et al. 2) and makes some extensions to this algorithm. The authors of [2] uses TextRank algorithm [39] for keyword selection and they select one third of words as keywords. In our implementation, we do not use this keyword rate but determine the number of keywords experimentally. Additionally, our algorithm makes use of the title of a documentary video in addition to the subtitles of the video. The title of a documentary video gives important clues about the video type because generally documentary names are selected in order to reflect the content of the documentary. For example; the category of the documentary “War of the Century” is “War”, or the category of the documentary “Planet Earth - Mountains” is “Geography”.

In our second categorization method, we perform categorization by using a learning module. This learning module learns the general WordNet domain distributions of categories. When categorizing a video, its WordNet domain

distribution is analyzed and the most similar category is assigned to the video.

For video summarization, we propose a novel approach by applying the text summarization techniques to video domain. We make use of two text summarization algorithms (Mihalcea and Tarau [39], Ercan and Cicekli [63]) and combine the results of these two algorithms to constitute a summary.

In this thesis, we combine the implementation of video categorization and video summarization in a system. Our system categorizes a video and presents a semantic summary of video automatically. Hence we provide users quick semantic information about video.

1.4 Thesis Outline

The rest of the thesis is organized as follows. In Chapter 2, we discuss the related work in video categorization and video summarization. For video summarization, we present the related work in both video summarization and text summarization. Chapter 3 describes our algorithms for video categorization and presents an evaluation of the algorithms. In Chapter 4, we give the description of our video summarization approaches and an evaluation of these approaches. Finally in Chapter 5, conclusions and possible future works are listed.

CHAPTER 2

RELATED WORK

In this chapter, we discuss the related work in video categorization and video summarization. In video categorization literature, there are three main approaches: visual-based approaches, audio-based approaches and text-based approaches. These approaches are presented in the subsections of video categorization related work section. In our thesis, we performed video summarization by utilizing text summarization techniques. Therefore, we present a summary of the literature on both video summarization and text summarization.

2.1 Related Work in Video Categorization

Today there are a lot of videos that people can access and finding a video of interest becomes a difficult job. Narrowing down the user's choices by categorizing videos can help people to solve this problem. Since there is a huge amount of videos to categorize, automatic video categorization is an important research area.

Video categorization algorithms assign a meaningful label to a video such as "sports video" or "comedy video". The required video features are drawn from three modalities: visual, audio and text. So, video categorization approaches can be classified as visual-based approaches, audio-based

approaches and text-based approaches. Also some approaches use a combination of visual, audio and text features.

2.1.1 Visual-Based Approaches

Generally, the approaches that utilize visual features extract features on a per frame or per shot basis. Most of the approaches to video classification rely on visual features, either alone or in combination with text or audio features.

A video is a collection of images and these images are called ***frames***. A ***shot*** is composed of frames within a single camera action. A ***scene*** is defined as one or more shots that form a semantic unit.

Visual features can be classified as Shot-Based Features, Object-Based Features, Motion-Based Features, Color-Based Features and MPEG Features.

Shot-Based Features: Many visual-based approaches use shots because a shot is a natural way to segment a video and shots generally represent a concept to humans such as "some people talking". A shot can be represented by a single frame, known as the key frame. Typically the first frame of a shot is selected as the key frame but some authors use the term, key frame, to refer to any single frame that represents a shot. One of the difficulties in using visual-based features is the large amount of potential data. This problem can be relieved by using key frames to represent shots.

Shots are closely related with some cinematic principles and give clues about cinema types. For example, action cinemas have shorter shots than character development cinemas. Or average shot length in music videos and commercials is shorter than sports. The length of shots is used in video classification in [9, 19, 22, 23, 24, 25, 26, 27].

To make use of shots, they must be detected first. Automatic detection of shots is a difficult job because there are various ways of making transition from one shot to the next. Shot transitions, mostly fall into one of the following categories: hard cuts, fades, and dissolves. In hard cuts, one shot stops suddenly and another begins. Fades are identified by slower changes in image features and they can be fade-out or fade-in. A fade-out is composed of a shot gradually fading out of existence to a monochrome frame. A fade-in occurs when a shot gradually fades into existence from a monochrome frame. If one shot fades in while another fades out, a dissolve occurs.

Shot transition types can be useful in features for classification. For example, shot transition using fades more often in commercials than they do in sports or news. The systems in [10, 24, 25, 26] use shot transitions in video classification.

Object-Based Features: Using object-based features is uncommon because detecting and identifying objects is difficult. When they are used, generally specific types of objects such as faces are identified [8, 9]. Besides this, some try to identify text objects within video frames [10, 11]. These text objects may be objects to be tracked or some character recognition methods can be applied to them.

Motion-Based Features: Motion within a video can be in two types: movement of the objects being filmed or movement due to camera actions. In some types of videos, there can be other types of movement. For example, in a news program, text can scroll at the bottom of the video. Motion-based methods mostly use MPEG motion vectors or the calculation of optical flow.

Optical flow is the pattern of motion in a sequence of images. It can be due to object motion or camera motion and it is calculated from the velocities of

pixel brightness patterns. The systems in [18, 23, 24, 25, 26, 28] use optical flow features in video classification.

The motion of foreground objects can be detected by using a frame-differencing approach. By using the Euclidean distance between pixels in the RGB color space, pixel-wise frame differencing of consecutive frames can be performed. The systems in [23, 26, 28, 29, 30] use frame-differencing feature in video classification.

***Color-Based Features*:** A video frame is composed of pixels and the color of each pixel is represented by a set of values from a color space. Two of the most popular color spaces are the red-green-blue (RGB) and hue-saturation-value (HSV) color spaces. The color distribution in a video frame is often represented by using a color histogram which shows the pixel count of each possible color in the frame. Similar frames have similar color histograms so color histograms are often used for comparing two frames. The authors of [7, 9, 15, 17, 18, 19, 20, 21, 22] use such color-based features in their classification system.

MPEG Features: One of the popular video formats is the Motion Pictures Expert Group (MPEG). For video classification, the main features that are extracted from MPEG videos are the Discrete Cosine Transform (DCT) coefficients and motion vectors. The usage of these features can improve the performance of the classification because these features have already been calculated and can be extracted without decoding the video. In [4, 12, 13] DCT coefficients are used and [8, 14, 15, 16, 17] use motion vectors in video classification.

2.1.2 Audio-Based Approaches

Audio-based approaches are slightly more common than text-based approaches in the literature. They typically require fewer computational resources than visual-based approaches. Another advantage of audio features is that if they need to be stored, they require less space.

Audio features can be extracted from either the time domain or the frequency domain. In time domain, the amplitude of a signal is plotted with respect to time. A signal in the time domain can be transformed to the frequency domain by using the Fourier transform.

Time-Domain Features: Whereas visual-based approaches try to use cinematic principles, the audio-based approaches try to approximate human perception of sound. The root mean square (RMS) of the signal energy approximates the human perception of the loudness or volume of a sound [31]. In [22, 33, 34] the RMS feature is used for classification.

The audio signal may be subdivided into sub bands and the energy of each sub band may be measured separately. Different classes of sounds fall into different sub bands [32].

The zero crossing rate (ZCR) is the rate of sign-changes along a signal in the current frame. Speech has a higher variability of the ZCR than in music. The systems described in [8, 12, 32] use zero crossing rate feature in classification.

Frequency-Domain Features: The frequency centroid is the midpoint of the spectral energy distribution. It approximates brightness and provides a measure of where the frequency components are concentrated [30]. Brightness in music is normally higher than in speech. The systems described in [16, 35, 36] use the frequency centroid feature in classification.

Bandwidth is a measure of the frequency range of a signal [31]. Some types of sounds have more narrow bandwidths than others. For example, speech has a lower bandwidth than music. In [8, 12, 16, 35, 36] bandwidth feature is used in classification.

Pitch is a measure approximated by the lowest frequency in a sample. It can be used to distinguish between male and female speakers. It can also be used to identify significant parts of a person's speech, (e.g. start of a new topic) [37].If a frame is not silent but does not have a pitch, it may represent noise or unvoiced speech [35].

2.1.3 Text-Based Approaches

There is not much research on text-based approaches and the main topic of this thesis is the categorization of videos using text features. There are some benefits of using text features in video categorization. First of all, text processing is simpler than video and audio processing and a lot of research about text categorization can be found in computational linguistics literature [81, 82, 83, 84, 85]. Also, the human language in a video carries more semantic information than its visual/audio features.

Words have meaning to humans and some words tend to be associated with certain categories. Another benefit of using text features is that by using some lexicon such as WordNet [5], concept learning can be performed.

However, using text features has also some disadvantages. Firstly, sometimes the transcript of video consists of a dialog and it does not describe what is seen on the screen. Secondly, text obtained by speech recognition or OCR of on-screen text may contain errors. And thirdly, not all videos can have text such as closed captions or subtitles.

The text associated with a video can be viewable text or transcript of the dialog. ***Viewable text*** is the text placed on the screen and some optical character recognition (OCR) methods should be used in order to use this text. The ***transcript*** of the dialog can be provided in the form of closed captions, open captions or subtitles. Alternatively, it can be obtained by using speech recognition methods.

Closed Captioned Text: Closed captions are the text displayed on video screen which provides transcription of the audio and some non-speech elements giving interpretive information to viewers such as sound effects (e.g., [BEAR GROWLS]), onomatopoeias (e.g., grrrr), and music lyrics (enclosed in music note symbols). Closed captioning lets hearing-impaired people to know what is being said in a video. Closed captions require a decoder to be seen on the television screen and it is possible to turn them on and off. Since closed captions are not part of the video, they can be extracted from the transmission of the video.

Zhu et al. [3] performs automatic news video story categorization based on the closed captioned text. They segment news video into stories using the demarcations which indicate the topic changes in the text. Then for each story, a category is assigned by extracting a list of keywords and further processing these keywords.

Brezeale and Cook [4] use text and visual features separately in video classification. As text, they use closed captions. To classify a movie, firstly the closed captions are extracted from the movie and stop words are removed from the closed captions. Then each word is stemmed by removing the suffixes to find its root. By using these stemmed words, a term-feature vector is generated. Classification is performed using support vector machine

(SVM) 6. There are 15 genres of movies from entertainment domain and the evaluation is performed on 81 movies. They state that the classification accuracy approaches 90%.

Speech Recognition: The transcript of the dialog can be extracted from speech using speech recognition methods. But the text derived from speech recognition generally has fairly high error rates. Wang et al. [8] makes classification by primarily using text features. They classify news video into one of ten categories. The spoken text is extracted using speech recognition methods.

OCR: By using some optical character recognition (OCR) methods, the text placed on the screen can be obtained and used. Qi et al. [38] classify a news video into types of news stories. The shots and if necessary scenes of video are firstly detected by using audio and visual features. Then the closed captions and scene text detected by the OCR methods are used in the classification of the news stories.

Subtitles: Subtitles can be defined as the textual version of a video's dialog or speech. Subtitles are prepared for people who can hear but can not understand the video because it is in another language.

Katsiouli et al. [2] uses subtitles for documentary classification. They perform categorization by using WordNet lexical database and WordNet Domains [5] and by applying natural language processing techniques on subtitles. They predefine documentary categories as Geography, History, Animals, Politics, Religion, Sports, Music, Accidents, Art, Science, Transportation, Technology, People and War. They state that their categorization approach has achieved 69.4% classification accuracy. In this thesis, we use a similar approach with different categorization algorithm, and the better results are obtained.

2.2 Related Work in Video Summarization

2.2.1 Video Summarization

The availability of digital video is increasing day by day at an exponential rate so users require assistance in accessing preferred video. Video summarization helps users to meet these needs by creating the condensed versions of videos by identifying the most important parts of a video.

Video summarization techniques analyze the content of a video and produce summaries by condensing this content into shortened forms. Videos have a multimodal nature and consist of multiple modes, such as sound, music, images and text. This makes summarization of video much more complex than summarization of text. In literature, there are approaches using the image features, audio features or text features in video summarization. Also some approaches use a combination of these features [1].

Image features include changes in color, texture shape and motion of objects generated by the image stream of the video. By using image features, the shots of a video can be identified, such as cuts or fades. Cuts are represented by sharp changes while fades are identified by slower changes in image features. For example, Ekin et al. [64] observed that the important scenes of a soccer game conforms to long, medium and close-up view shots and they use these shot types in their summarization system.

In addition to shots, specific objects and events can be identified by analyzing image features and this information could improve summarization performance. Knowledge of content domain could be helpful in the identification of objects within the video (e.g. anchor person) and events (e.g. the news headlines). For example, news video normally starts with an

overview of headlines, continue with a series of reports and end in a return to the anchor person. Such domain analysis improves summarization performance. The systems in [64, 65, 66, 67, 68, 69] use image features to identify representative key frames for inclusion in the video summary and all are non-domain specific. The techniques presented in [70, 71, 72, 73] analyze image features from the video stream, but they are domain specific.

Audio features associated with a video include speech, music, sounds and silence. These features are used for selecting candidate segments to be included in a video summary and domain specific knowledge can be used to enhance the summary success. For example, excited commentator speech and excited audience sounds may show a number of potential events such as the start of a free kick, penalty kick, foul, goal etc. [74]. Rui et al. [75] analyze the speech track to find exciting segments and events such as baseball hits in baseball videos.

Text features associated with a video can be viewable text placed on the screen or transcript of the dialog which can be provided in the form of closed captions, open captions or subtitles. Text features plays an important role in video summarization as it contains detailed information about the video content. Pickering et al. [77] make summarization of television news by using the accompanying subtitles. They extract news stories from the video and provide a summary for each story by using lexical chain analysis. Tsoneva et al. [78] creates automatic summaries for narrative videos using textual cues available in subtitles and scripts. They extract features like keywords, main characters names and presence, and according to these features they identify the most relevant moments of video for preserving the story line.

2.2.2 Text Summarization

Text summarization can be defined as identifying the significant parts of a text to constitute a summary. Text summarization techniques can be useful in video summarization since some videos have text related with the content of the video and the summary of such a text could be an important resource in video summarization.

Text summarization techniques investigate different clues that could be used to identify important topics and ideas of the text. The summarization methods can be classified by the clues that they use in summarization.

Methods Using Position in Text

Some authors observed that important content of a text is usually positioned in the first sentences. As a result of this observation, a very simple and surprisingly successful method for summarization is emerged. The authors of [48, 49, 50, 51] created summaries by selecting the first sentences of text and they state that this simple technique gives well results in news articles and scientific reports.

Methods Using Cue-Phrases and Formatting

In text, to emphasize the importance of a sentence some phrases are used such as "significantly", "in conclusion" and these phrases are called ***bonus phrases***. On the other hand, some phrases reflect the unimportance of a sentence such as "hardly", "impossible" and these phrases are called ***stigma phrases***. In addition to cue-phrases, some formatting features like bold words, headers could enhance the summarization performance. The systems in [48] and [50] make use of cue phrases and format features in their summarization systems.

Methods Using Lexical Cohesion

Some authors use weighted vectors of TF*IDF (Term Frequency * Inverse Document Frequency) values to represent sentences. Tf*Idf value takes advantage of word repetition in a text which is a lexical cohesion type. Radev et al. [51] uses such weighted vectors to find the important sentences in summarization. Erkan [52] developed an algorithm similar to Google's Pagerank [53] for the selection of summary sentences. It is an important algorithm in this research area. Mihalcea et al. [39] proposed a summarization algorithm named TextRank. TextRank algorithm relies on the Google's Pagerank [39] algorithm and uses the word repetition feature.

Lexical chains, which are sets of related words, also can be used for modeling lexical cohesion. Barzilay [54] used lexical chains to extract summaries and achieved good results. Many lexical cohesion based algorithms are developed following the Barzilay's algorithm. Silber and McCoy [55] proposed a summarizer based on lexical chains and tried to improve the running time of lexical chaining algorithm. Brunn et al. [56] and [57] used lexical chains and offered a different sentence selection approach. Ercan and Cicekli [63] exploit the lexical cohesion structure of the text to determine the importance of sentences. Their summarization algorithm constructs the lexical chains of a text and identifies topics from lexical chains. The text is segmented with respect to these topics and the most important sentences are selected from these segments. Also [58, 59, 60] use lexical chains for summarization. Ye et al. [61] proposed a new approach by using WordNet and WordNet glosses for summarization.

CHAPTER 3

VIDEO CATEGORIZATION

Today people have access to a large amount of video so that finding a video of interest becomes difficult. One method that viewers use to narrow down their choices is to look for video within specific categories or genre. Since there is huge amount of video to categorize, the automatic categorization would be very useful.

The categorization of videos can be realized by using visual features, audio features or text features related to the video. Using text features have several benefits over visual/audio features. Firstly, text processing is a more lightweight process than video/audio processing. Also, there is an extensive research about text categorization in computational linguistics literature and this information can be used in video categorization. In addition, text of a video, such as subtitles, carries more semantic information than visual/audio features.

In this thesis, we propose two algorithms for video categorization: Category Label Assignment and Categorization by Learning. In our first algorithm, we reference the Katsiouli et al. [2] and we make some extensions to their approach. We add video name processing and we change the way of determining the number of keywords in subtitle processing. By these extensions, better results are obtained. According to this algorithm, a video is assigned a category label by using WordNet lexical database and WordNet

Domains [5] and by applying natural language processing techniques on its subtitles.

In our second algorithm, we make categorization by learning. We develop a learning module which can be trained with the videos with known categories. The algorithm starts with the preprocessing steps of the first algorithm and categorization is performed by the learning module.

This chapter is designed as follows: the common preprocessing steps of two algorithms are given in Section 3.1. The first video categorization algorithm is given in Section 3.2, the second video categorization algorithm is given in Section 3.3, and the evaluation of these algorithms is given in Section 3.4.

3.1 General Categorization Framework

The two video categorization algorithms of our system have some common steps. By these steps, the WordNet domains of a video are extracted and the extracted domains are used in the categorization algorithms. The overview of extracting WordNet domains is given in Figure 3.

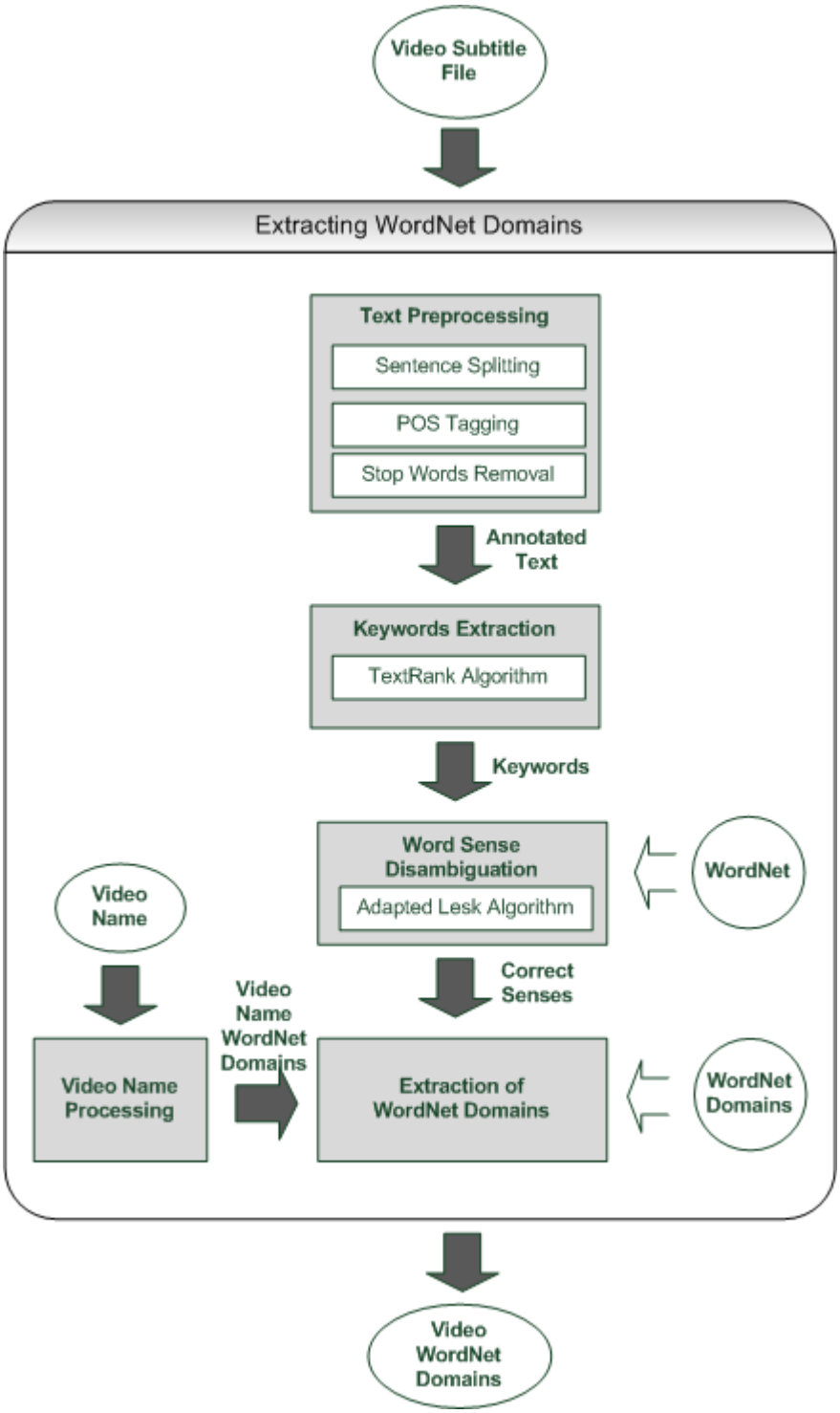

Figure 3 Extracting WordNet Domains

Extracting WordNet domains starts with “Text Preprocessing”. In this step, the sentences in the subtitle file are split, the words in every sentence are tagged with POS tags and the stop words are removed from the sentences. This processed text is given to “Keywords Extraction” module and this module finds the keywords of the given text. Since these keywords may carry more than one meaning, the “Word Sense Disambiguation” module finds the correct senses of the keywords by using an adaptation of the Lesk algorithm [40]. Then “WordNet Domains Extraction” module finds the WordNet domains of the keywords corresponding to the correct senses found. This module uses WordNet Domains and considers the effect of the video name on categorization. Since video titles give important clues about the video category, this information is taken into consideration. Hence we obtain the WordNet domains of the video. The details of these steps are given in the following subsections.

3.1.1 Text Preprocessing

In the text preprocessing step, the subtitle file is processed and sentences of the subtitle file are split from each other. A sample subtitle file is shown in Figure 4 and the sentences created from this subtitle file are shown in Figure 5.

Then a part of speech (POS) tagger is applied to the words of the sentences. Part of speech tagging is the process of marking up the words in a text as corresponding to a particular word class, based on both its definition, and its context, i.e., relationship with adjacent and related words in a phrase, sentence, or paragraph. A word can have different part of speech tags (such as noun, verb, adverb, adjective), but each word in the sentence can belong to exactly one word class.

1
00:00:25,600 --> 00:00:31,080
Human beings venture into the highest parts of our planet at their peril.
2
00:00:31,640 --> 00:00:34,480
Some might think that by climbing a great mountain
3
00:00:34,560 --> 00:00:36,320
they have somehow conquered it,
4
00:00:36,720 --> 00:00:39,800
but we can only be visitors here.
5
00:00:42,160 --> 00:00:46,240
This is a frozen alien world.

Figure 4 A sample subtitle file

Human beings venture into the highest parts of our planet at their peril.

Some might think that by climbing a great mountain they have somehow conquered it, but we can only be visitors here.

This is a frozen alien world.

Figure 5 The split sentences of the subtitle file

Human/**JJ** beings/**NNS** venture/**NN** into/**IN** the/**DT** highest/**JJS** parts/**NNS** of/**IN** our/**PRP$** planet/**NN** at/**IN** their/**PRP$** peril/**NN**.

Some/**DT** might/**MD** think/**VB** that/**IN** by/**IN** climbing/**VBG** a/**DT** great/**JJ** mountain/**NN** they/**PRP** have/**VBP** somehow/**RB** conquered/**VBN** it/**PRP**, but/**CC** we/**PRP** can/**MD** only/**RB** be/**VB** visitors/**NNS** here/**RB**.

This/**DT** is/**VBZ** a/**DT** frozen/**JJ** alien/**JJ** world/**NN**.

Figure 6 Tagged Sentences

A part of speech tagger determines the word class of each word in the sentence. The Stanford Log-linear Part-Of-Speech Tagger [41] is used for this purpose. The assigned part of speech tags consist of coded abbreviations conforming to the scheme of the Penn Treebank [42], the linguistic corpus developed by the University of Pennsylvania. The part-of-speech tags of the sentences in Figure 5 are shown in Figure 6. For example, "JJ" means "Adjective", "NNS" means "Noun, plural", and NN means "Noun, singular or mass". The whole list of POS tag abbreviations and corresponding Penn Treebank descriptions can be found in Appendix A.

After part of speech tagging, stop words are removed from sentences since these words carry no semantics. Stop words are words that do not contribute to the meaning of the sentence such as "above", "the", "her". The list of the used stop words is given in Appendix B.

3.1.2 Keywords Extraction

In order to select the most important words in the subtitle file for classifying the video, a keyword selection algorithm, namely the TextRank [39] algorithm, is used. TextRank is a well known algorithm among the text classification community and used for text applications such as keywords extraction and text summarization.

The TextRank algorithm builds a graph representing the text and applies a ranking algorithm to the vertices of the graph. For keywords extraction, the words of the text are added to the graph as vertices. Two vertices are connected if they have a *co-occurrence* relation. Two vertices *co-occur* if they are within a window of maximum N words, where N can be set from 2 to 10. In our implementation N is set to 2. A sample graph build for a part of "*Wildlife Specials – Tiger*" documentary subtitle is shown in Figure 7.

Most of us would agree that a tiger is one of the world's most beautiful creatures. Sadly, in the wild, it's threatened with extinction. But, fortunately, it breeds very well in captivity, as this little cub proves. But is a tiger in a cage truly a tiger? I doubt it. To see the true essence and beauty of a tiger, you have to see it in the wild. This is the story of a tigress in the heart of India. Our tigress lives in Kanha National Park.

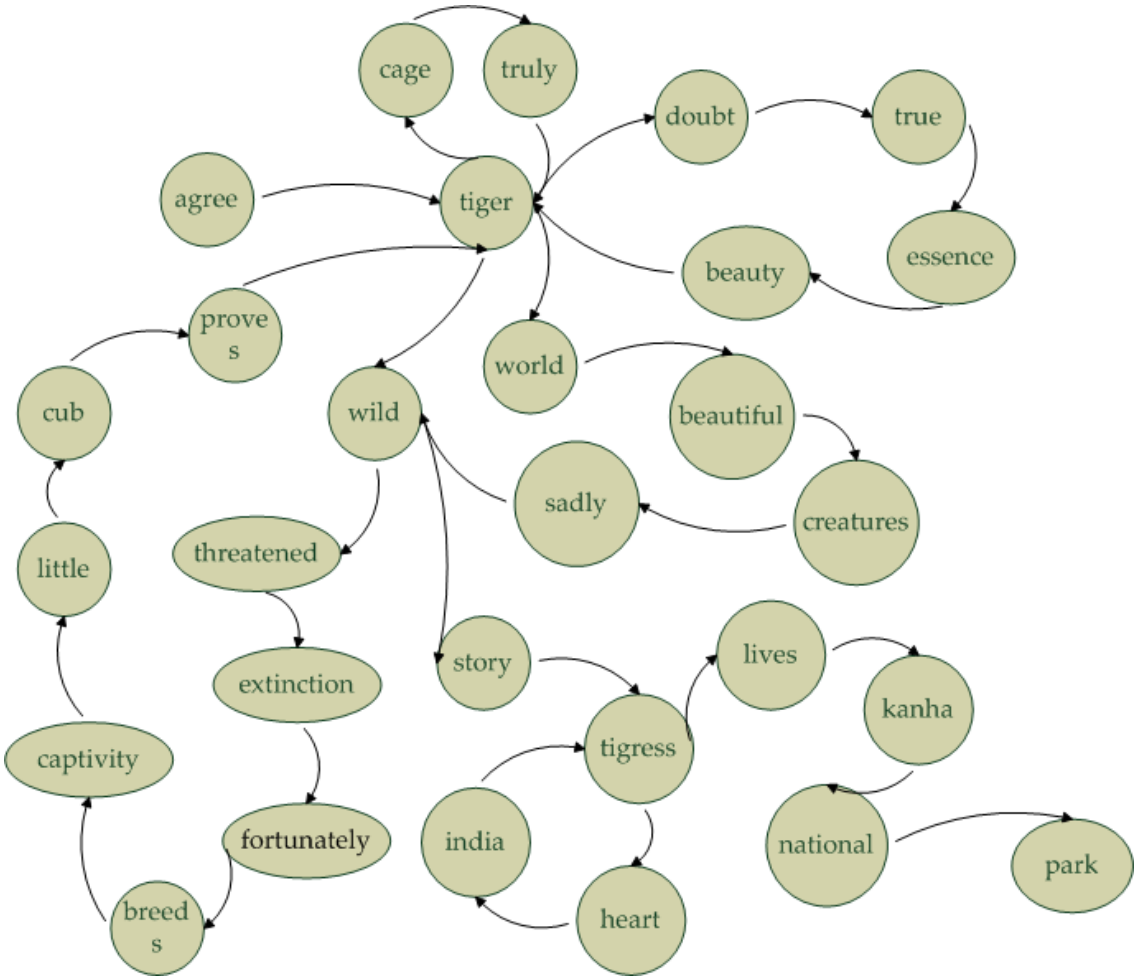

Figure 7 Sample graph build for a part of "*Wildlife Specials – Tiger*" documentary subtitle

Then for deciding the importance of a vertex, a graph-based ranking algorithm derived from Google's PageRank algorithm [44] is used. The basic idea of the algorithm is "voting": when a vertex links to another one, it casts a vote for that vertex. Also, the importance of the vertex casting the vote, determines the importance of the vote. Hence, the score of a vertex is computed by the votes that are cast for it and the score of the vertices casting these votes.

Formally, the score of a vertex V_i is defined as [39]:

$$S(V_i) = (1-d) + d * \sum_{j \in In(V_i)} \frac{1}{|Out(V_j)|} S(V_j) \qquad (3.1)$$

Here, $In(V_i)$ is the set of vertices that point to Vi and $Out(V_j)$ is the set of vertices that Vertex V_j points to. d is damping factor that can be set between 0 and 1. d is usually set to 0.85 and we also use this value in our implementation. After a certain number of iterations, the final values of all vertices are computed.

Once the score of each vertex is computed, the vertices are sorted based on their scores and top T vertices are selected as keywords. Generally, T is set to a third of the number of vertices in the graph. The keywords of the text in Figure 7 extracted by TextRank algorithm are shown in Figure 8.

Keywords Assigned by TextRank:
tiger, tigress, wild, national, kanha, captivity, breeds, lives, little

Figure 8 Keywords of the text in Figure 7

In our implementation, the number of the vertices selected as keyword is determined experimentally and the details are given in Section 3.4.

3.1.3 Word Sense Disambiguation

Most words in natural languages have multiple possible meanings or senses. For example, the word "bank" may mean "an institution for saving and borrowing money" or "a simple seat usually found in gardens". In the sentence "The bank down the street was robbed!", "bank" refers to an institution for saving and borrowing money. Humans can easily understand which sense of a word is intended by using his experience of the world and language but the computer programs do not have such a benefit.

Word Sense Disambiguation (WSD) is the task of determining the correct sense of a word in a text. In order to find the correct senses of the keywords, we applied a WSD algorithm, which is presented in [40]. This algorithm is an adaptation of Lesk's dictionary-based word sense disambiguation algorithm [43].

Lesk's algorithm disambiguates words in short phrases and uses the glosses found in traditional dictionaries. The gloss of each sense of a word in a phrase is compared to the glosses of other words in the phrase. The sense whose gloss shares the largest number of words in common with the glosses of the other words in the phrase is selected as the word's sense.

The adapted Lesk algorithm [40] uses WordNet to include the glosses of the words that are related to the word being disambiguated through semantic relations, such as hypernym, hyponym, holonym, meronym, troponym, and attribute of each word. This supplies a richer source of information and increases disambiguation accuracy.

WordNet is a dictionary but while traditional dictionaries are arranged alphabetically, WordNet is arranged semantically. Nouns, verbs, adjectives and adverbs are grouped together to form synonyms (synsets) which expresses a distinct concept. Synsets are connected to each other through some semantic relations. We give the definitions of the relations used in the adapted Lesk algorithm: hypernym, hyponym, holonym, meronym, troponym, and attribute.

Hypernym / hyponym is a generalization / specialization relation; if synset X is a kind of synset Y, then X is the hyponym of Y, and Y is the hypernym of X. For example; "vegetables" is a hypernym of "broccoli", and conversely, "broccoli" is a hyponym of "vegetables".

Holonym / meronym is a whole of / part of relation; if synset X is a part of synset Y, then X is a meronym of Y. Conversely, if synset Y has synset X as a part, then Y is a holonym of X. For example; "building" is a holonym of "window", and conversely, "window" is a meronym of "building". ***Hypernym / troponym*** is a relation between verbs; synset X is the hypernym of Y, if Y is one way to X; Y is then the troponym of X. For example; "to lisp" is a troponym of "to talk", and conversely, "to talk" is a hypernym of "to lisp".

Attribute is a relation between an adjective and a noun; for example, the attribute of adjective "beautiful" is the noun "beauty".

The adapted Lesk algorithm compares glosses between each pair of words in the window of context. These glosses are the glosses associated with the synset, hypernym, hyponym, holonym, meronym, troponym, and attribute of each word. For example, the gloss of a synset of one word can be compared with the gloss of a hypernym of the other word.

When comparing two glosses, a score is computed for them. To compute the score, an "overlap" parameter is used. "Overlap" is the longest sequence of one or more consecutive words that occurs in both glosses. The overlaps which are made up from entirely non-content words, such as pronouns, prepositions, articles and conjunctions, are ignored. The square of the number of words in the overlaps are added, hence a score is computed.

Once all the gloss comparisons have been made, the candidate combination with the highest score is selected and the target word is assigned the sense in that combination.

In our algorithm, word sense disambiguation is essential for finding the WordNet domains of the words. Since, we try to find the WordNet domains of keywords in the next step; we need to find the correct senses of these words. In our implementation, by using the adapted Lesk algorithm, the correct senses of the keywords are assigned. For the keywords in Figure 8, the senses assigned by the adapted Lesk algorithm are given in Table 1.

3.1.4 Extraction of WordNet Domains

By augmenting WordNet with domain labels, **WordNet Domains** were created [5]. The synsets in WordNet have been annotated with at least one domain label by using a set of about two hundred labels hierarchically organized. If there is no appropriate domain label for a synset, the label "factotum" was assigned to it. In Table 2, the senses of the word "*bank*" and the corresponding WordNet domains are shown; the example is taken from [45].

Table 1 Senses of the keywords in Figure 8

Word	Pos Tag	Sense	Synset (Gloss)
tiger	Noun	1	tiger, Panthera tigris -- (large feline of forests in most of Asia having a tawny coat with black stripes; endangered)
tigress	Noun	0	tigress -- (a female tiger)
wild	Adjective	1	wild, untamed -- (in a natural state; not tamed or domesticated or cultivated; "wild geese"; "edible wild plants")
national	Noun	0	national, subject -- (a person who owes allegiance to that nation; "a monarch has a duty to his subjects")
kanha	Noun	-1	Not Found In WordNet Dictionary
captivity	Noun	0	captivity, imprisonment, incarceration, immurement -- (the state of being imprisoned; "he was held in captivity until he died"; "the imprisonment of captured soldiers"; "his ignominious incarceration in the local jail"; "he practiced the immurement of his enemies in the castle dungeon")
breeds	Verb	3	breed, multiply -- (have young (animals); "pandas rarely breed in captivity")
lives	Verb	0	dwell, shack, reside, live, inhabit, people, populate, domicile, domiciliate -- (make one's home or live in; "She resides officially in Iceland"; "I live in a 200-year old house"; "These people inhabited all the islands that are now deserted"; "The plains are sparsely populated")
little	Adjective	3	little, small -- (not fully grown; "what a big little boy you are"; "small children")

Table 2 Senses of the word "bank" with their corresponding WordNet domains

Sense Number	Synset (Gloss)	Domains
1	depository financial institution, bank, banking concern, banking company (a financial institution ...)	**Economy**
2	bank (sloping land ...)	**Geography, Geology**
3	bank (a supply or stock held in reserve...)	**Economy**
4	bank, bank building (a building...)	**Architecture, Economy**
5	bank (an arrangement of similar objects...)	**Factotum**
6	savings bank, coin bank, money box, bank (a container...)	**Economy**
7	bank (a long ridge or pile...)	**Geography, Geology**
8	bank (the funds held by a gambling house...)	**Economy, Play**
9	bank, cant, camber (a slope in the turn of a road...)	**Architecture**
10	bank (a flight maneuver...)	**Transport**

Table 3 WordNet domains of the words in Table 1

Word	Sense	WordNet Domains
tiger	1	animals, biology
tigress	0	animals
wild	1	factotum
national	0	politics
kanha	-1	-
captivity	0	factotum
breed	3	factotum
live	0	town_planning
little	3	factotum

In this step, we find the WordNet domains of the keywords. In finding the domains of a word, we should know the synset (gloss) of that word. Since we have found the synsets of keywords in the previous step, we make use of this information in finding the WordNet domains. The WordNet domains of the words in Table 1 are given in Table 3.

Afterwards, we calculate the occurrence score of each domain label. That is, how many times a domain label appears in the keywords' domains. Then we sort these domains in decreasing order.

3.1.5 Video Title Processing

We observed that video titles give important clues about video categories. For example, the category of the documentary "War of the Century" is "War", or the category of the documentary "Art of the Spain" is "Art". So, as an extension to the approach of Katsiouli et al. [2], we decided to make use of the video name information when categorizing the video. Utilizing the video title in video categorization has increased the performance of our categorization algorithms.

For this purpose, the WordNet domains are found for each word in the video title. Hence a list of WordNet domains which describes the video title is acquired. For example, the WordNet domains of the video title, "Wildlife Specials - Tiger", in Figure 7 are "animals", "biology" and "factotum".

In the previous step, we have obtained WordNet domains of the video keywords and the occurrence scores of these domains. If one of these domains also exists in the video title domains, the occurrence score of the domain is increased by the ratio of one fourth. This ratio is determined experimentally and the details of the experiments are given in Section 3.4.

At the end of this step, we obtain the WordNet domains of a video with their occurrence scores.

3.2 Category Label Assignment

Our first video categorization algorithm is "Category Label Assignment" by using "Mappings Between Categories and WordNet Domains". In this algorithm we took the approach of Katsiouli et al. [2] as a basis, but we make some differences in implementing the steps and by these differences we get better results. These differences are given in detail in Section 0.

In this video categorization algorithm, we find the WordNet domains related to the categories and a category label is assigned to the video by comparing the video domains with category domains. The overview of the algorithm is given in Figure 9 and the details of these steps are given in the following subsections.

3.2.1 Defining Video Categories

In this thesis, we have chosen documentary videos as the categorization domain. It is easier to classify documentaries since they are generally restricted to a specific domain. 14 documentary categories are defined, namely, Geography, History, Animals, Politics, Religion, Sports, Music, Accidents, Art, Science, Transportation, Technology, People and War.

In order to assign a category label to documentary videos, a mapping is defined between category labels and WordNet domains. First, the senses related to each category label are acquired from WordNet. Also the senses related with category label through hypernym & hyponym relations are collected.

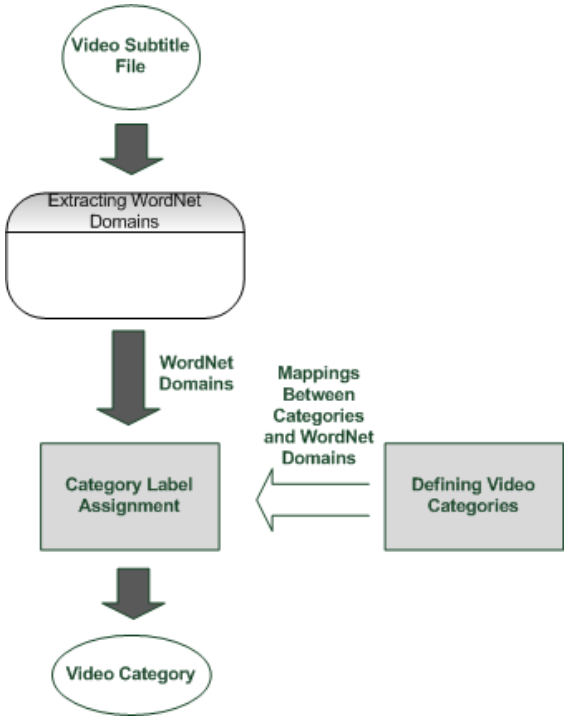

Figure 9 Category Label Assignment

Table 4 Category labels and corresponding WordNet domains

Category	**Top Rank WordNet Domains**
Geography	geography
Animals	animals, biology, entomology
Politics	politics, psychology
History	history, time_period
Religion	religion
Transportation	transport, commerce, enterprise
Accidents	transport, nautical
Sports	sport, play, swimming
War	military, history
Science	medicine, biology, mathematics
Music	music, linguistics, literature
Art	art, painting, graphic_arts
Technology	engineering, industry, computer_science
People	sociology, person

Then the WordNet domains that correspond to the senses of each category label are obtained. Afterwards, for each category, the occurrence score of derived domains are calculated and sorted in decreasing occurrence order. Table 4 shows the category labels and corresponding top-ranked WordNet domains that are determined by Katsiouli et al. [2].

3.2.2 Category Label Assignment

In this step, a category label is assigned to the video. For this purpose, the sorted WordNet domains of the video are compared to the top-rank domains of the categories.

The algorithm compares the first domain of the video with the first domains of the categories.

- If the first domain of a category is equal to the first domain of the video, this category label is assigned to the video.
- If the first domain of more than one category is equal to the first domain of the video, the second domain of the corresponding sets are compared, and so on.
- If none of the category's first domain is equal to the first domain of the video, then the second domain of the video is compared to the first domains of the categories.

The algorithm continues as described above until a category label is assigned to the video. For example; let the following table shows the category labels and corresponding WordNet domains;

Table 5 Sample category labels and corresponding WordNet domains

Category	Top Rank WordNet Domains
Animals	animals, biology, entomology
Transportation	transport, commerce, enterprise
Accidents	transport, nautical

If the sorted WordNet domains of a video are:

- "animals, entomology, biology", then it is assigned "Animals" category.
- "transport, nautical, geography", then it is assigned "Accidents" category.
- "geography, animals", then it is assigned "Animals" category.

In this video categorization algorithm, we referenced the Katsiouli et al. [2] approach but by making some changes in implementation, we get better results. At the text preprocessing step, while they use the Mark Hepple's pos tagger [86], we use the Stanford Log-linear Part-Of-Speech Tagger [41] for that purpose. Then in keyword extraction phase, the authors of [2] use a third of the number of words as the keyword count. In our system, we determined this number experimentally and changing the number of keywords affected the system's classification accuracy. Also in our implementation, we considered the effect of the video title since video titles give strong clues about the categories of documentary videos.

First of all, we implemented the approach of Katsiouli et al. [2] and evaluated with 40 documentary subtitles from National Geographic and BBC. In this situation, we get %60 classification accuracy. After the changes we made to the algorithm, we get %75 classification accuracy on the same experiment set.

3.3 Categorization by Learning

Our second video categorization algorithm named as "Categorization by Learning" uses a "Learned Category/Domain Distribution" matrix. In this algorithm, we propose a learning mechanism to assign video a category label. In the preprocessing steps of the algorithm, we reference the approach of Katsiouli et al. [2].

This algorithm includes a learning phase where the domain distributions of categories are learned from videos with known categories. When a video is to be categorized, the domain distribution of the video is compared with the learned domain distributions of categories. The most similar category is assigned to the video. The overview of the algorithm is given in Figure 10 and the details of these steps are given in the following subsections.

3.3.1 Learning Category Domain Distribution

In this phase, the domain distribution of categories is learned. For this purpose, documentaries with known categories are used. When selecting documentaries for learning purpose, it is important to select documentaries from all category labels. For each previously defined category, the learning is realized as described in the following.

First of all, the documentary subtitles belonging to a specific category are processed by using the "Extracting WordNet Domains" module discussed in Section 3.1. Hence, the domains and domain occurrence scores of the category are collected.

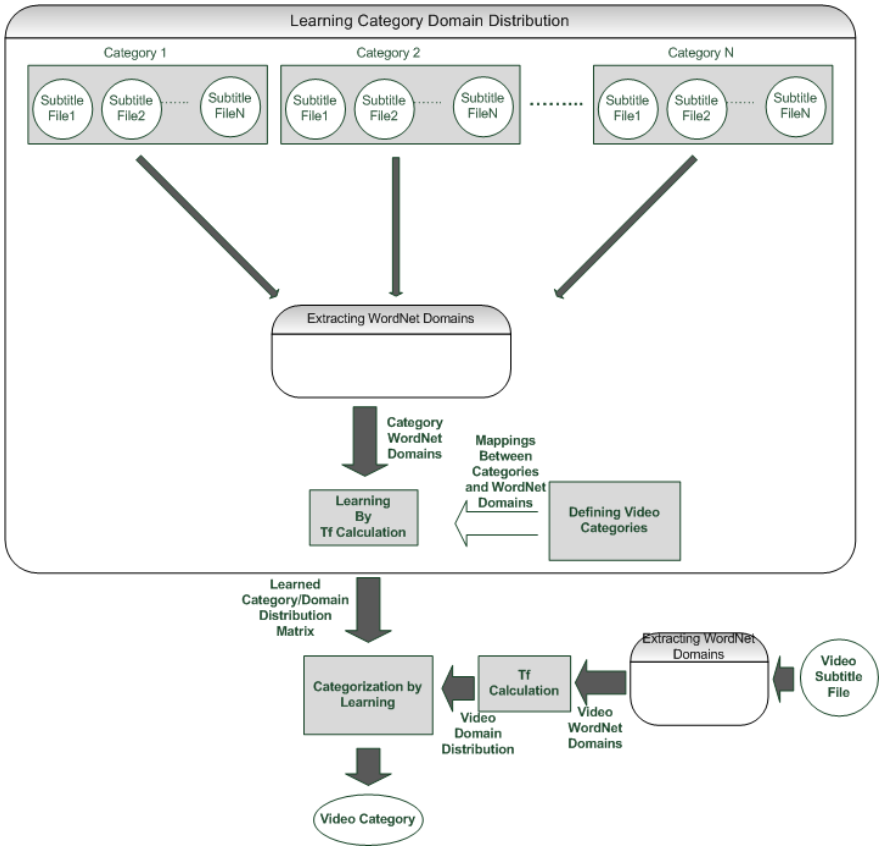

Figure 10 Categorization by learning

In order to determine the domain distribution of the category, we first intended to use Tf*Idf value of each domain. The ***Tf*Idf (term frequency–inverse document frequency)*** [46] is a weight generally used in information retrieval and text mining. The ***Tf (term frequency)*** is a measure of the importance of the term t_i within the particular document d_j and defined as follows;

$$TF_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \tag{3.2}$$

Here, $n_{i,j}$ is the number of occurrences of the considered term (t_i) in document d_j, and the denominator is the sum of number of occurrences of all terms in document d_j.

The ***Idf (inverse document frequency)*** is a measure of the general importance of a term t_i and defined as follows;

$$IDF_i = \log \frac{|D|}{|\{d : t_i \in d\}|} \tag{3.3}$$

Here, $|D|$ is the total number of documents in the corpus and $|\{d : t_i \in d\}|$ is the number of documents where the term t_i appears.

Then;

$$(TF*IDF)_{i,j} = TF_{i,j} * IDF_{i,j} \tag{3.4}$$

In our implementation, we wanted to use TF*IDF values computed for domains and categories but we observed that sometimes the number of categories where a domain appears is equal to the total number of categories. In this case, the IDF value of this domain becomes zero hence also the TF*IDF

value becomes zero. We do not want to have such an effect because we want to obtain distribution of all domains in a category. So instead of using TF*IDF weight, we have used TF weight of domains to determine the domain distribution of categories.

The TF weight is computed for each category and domain pair. Hence a matrix showing the domain TF weights of all categories is obtained.

As an example, 22 documentaries from different categories are used for learning. The documentaries used for learning are shown in Table 6.

Table 6 Documentaries used for learning

Documentary Name	**Category**
Charles Lindbergh - The Lone Eagle	People
Code of the Maya Kings	Geography
Islam Empire of Faith - 01	Religion
Islam Empire of Faith - 02	Religion
Lost Ships of the Mediterrian	Geography
Mysteries of Egypt	History
Pearl Harbor Legacy of Attack - 01	War
Planet Earth - From Pole to Pole	Geography
Planet Earth - Mountains	Geography
The Art of Spain - The Dark Heart	Art
The Art of Spain - The Moorish South	Art
The Battle for Midway	War
The Incredible Human Body	Science
The Pink Floyd Story	Music
The Power of Nightmares - The Phantom Victory	Politics
The Power of Nightmares - The Shadows in the Cave	Politics
The Secret Life of Cats	Animals
Those Wonderful Dogs	Animals
Tropicalia Revolution	Music
War of the Century - 01	War
War of the Century - 02	War
Wildlife Specials - Leopard	Animals

Table 7 shows a sample part of the computed matrix representing the TF values of category domain pairs. The complete matrix is given in Appendix D.

Table 7 Sample part of the matrix representing the domain distribution of categories

	Geography	**Animals**	**Politics**
geography	0,0552444	0,032574	0,039201
animals	0,0302953	0,053447	0,009224
biology	0,0315682	0,041429	0,016141
entomology	0,0022912	0,000949	0
politics	0,0068737	0,008223	0,037663
psychology	0,0043279	0,007906	0,008455
history	0,0099287	0,008223	0,01691
time_period	0,0313136	0,023087	0,017294
religion	0,0089104	0,004744	0,021522
transport	0,0129837	0,012334	0,013451

3.3.2 Categorization by Learning

In this step, the learned information is used for categorization. In the previous step, we have learned the domain distribution of the categories. When we want to categorize a video, we compare this video's domain distribution with the domain distribution of categories and we select the category which has the most similar domain distribution with the video. The categorization of a documentary video is performed as follows.

The subtitle of the video is processed by using the first four steps of our categorization algorithm. Hence, the domains and domain occurrence scores of the video are obtained.

Then, in order to determine the domain distribution of the video, TF value of each domain is computed as described in the previous step.

After this, by using the matrix we have obtained in the previous step, we try to find the category which has the most similar domain distribution with the video. For this purpose, we used the cosine similarity. ***Cosine similarity*** [47] is a measure of similarity between two vectors by finding the cosine of the angle between them. This value is often used to compare documents.

The cosine similarity of two vectors, A and B, is represented using a dot product and magnitude as;

$$similarity = \cos(\theta) = \frac{A.B}{\|A\| \|B\|} \quad (3.5)$$

The ***dot product***, also known as the scalar product, is an operation which takes two vectors and returns a scalar quantity. The ***magnitude*** of a mathematical object is its size; in cosine similarity formula, it is the length of the vector.

For example, we want to categorize a documentary video named "The Everest". So we compute the domain distribution of the video, Table 8 shows the domain distribution of the documentary "The Everest".

Then by using the matrix learned in the previous step, we compute the similarities of categories. Table 9 shows the cosine similarities between the documentary "The Everest" and the categories. Since "Geography" category is most similar to the "The Everest" documentary, it is assigned as the documentary category.

Table 8 Domain distribution of the documentary "The Everest"

Domain	Tf Value
geography	0,036053131
animals	0,024667932
biology	0,032258065
entomology	0
politics	0,009487666
psychology	0,004743833
history	0,006641366
time_period	0,030360531
religion	0,011385199
transport	0,018975332
commerce	0,0028463
enterprise	0
nautical	0,003795066
sport	0,010436433
play	0,0056926
swimming	0
military	0,008538899
medicine	0,012333966
mathematics	0,0028463
music	0,0056926
linguistics	0,004743833
literature	0,004743833
art	0,003795066
painting	0
graphic_arts	0
engineering	0,000948767
industry	0
computer_science	0,003795066
sociology	0,0056926
person	0,025616698

Table 9 Cosine similarities between the documentary "The Everest" and the categories

Category	Cosine Similarity
Geography	0,9590928
History	0,9452289
People	0,9376402
Animals	0,9267696
Science	0,9037822
Music	0,871545
Religion	0,825404
Politics	0,8154419
War	0,8115139
Art	0,7872766

3.4 Experiments and Evaluation

In order to evaluate the effectiveness of our categorization algorithm, we used documentaries from BBC and National Geographic. The list of documentaries used in evaluation and their original categories are given in Table 13 in APPENDIX C.

We performed the evaluation using the Classification Accuracy (CA) metric. ***Classification Accuracy (CA)*** reflects the proportion of the program's correct assignments that agree with the original assignment.

For our first categorization algorithm, "Category Label Assignment", we conduct several experiments by changing some of the parameters. First of all, for keyword extraction, we change the number of keywords selected and observe the results. In the algorithm we referenced (Katsiouli et al. [2]), a third of the words are selected as keywords. So we start with trying this rate and then by changing this rate we want to experiment the effect of the rate of keywords. Figure 11 shows the classification accuracy (CA) with changing keyword rates.

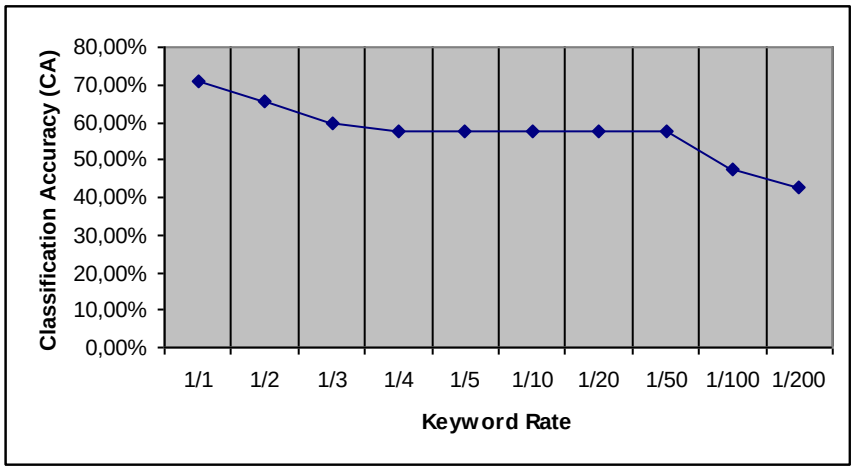

Figure 11 Classification Accuracy with Keyword Rates

For example, if we select the "1/3" of the words as keywords, we get "60%" classification accuracy. We observe from the Figure 11 that increasing the rate of the keywords improves the classification accuracy.

In addition to the *keyword rate* parameter, we observe that in TextRank algorithm [39] all words are assigned a weight and selecting words above a certain weight could be an alternative for determining the number of keywords. Therefore, words above a certain weight are selected as keywords and the classification accuracy of the system is computed for changing weights. The diagram below shows the classification accuracy (CA) with changing keyword weights.

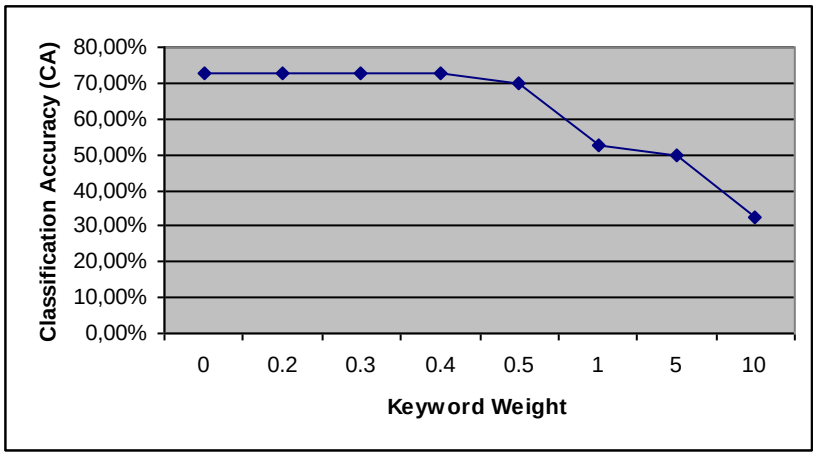

Figure 12 Classification Accuracy with Keyword Weights

For example, if we select the words with weight bigger than "5" as keywords, we get "50%" classification accuracy. As seen from the Figure 12, we get best results when using the weights between 0 and 0.4. Therefore using any weight between 0 and 0.4 does not change the CA, but selecting higher weights decreases the number of keywords. Using less number of keywords decreases the computation time. Therefore the upper bound value "*weight > 0.4*" could be preferred to the others. So we use the experimentally determined value "*weight > 0.4*"as the keyword selection parameter in our video categorization algorithm.

Afterwards we make some experiments by considering the effect of video title when categorizing the video. In the algorithm, we extract the WordNet domains of a video and the occurrence scores of these domains. If one of these domains also exists in the video title domains, the occurrence score of the domain is increased by some ratio. The diagram below shows the affect of this ratio on the performance of the video categorization system.

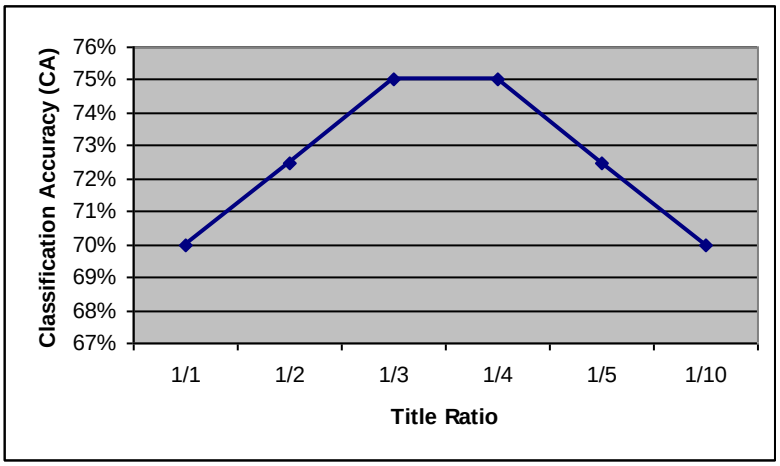

Figure 13 Classification Accuracy with Title Ratio

When we increase the occurrence score of domains which also exist in the video title domains by the ratio of "*one third*" or "*one fourth*", we get a classification accuracy of %75 as the best result. Selecting "*one third*" or "*one fourth*" does not make a significant difference in computation time, so any of them could be used in video categorization algorithm and we select "*one fourth"* in our implementation.

Our second video categorization algorithm is "Categorization by Learning". In the WordNet domains extracting part of this algorithm, we use the parameters which we get the best results in the experiments of our first video categorization algorithm. Namely, we use the keywords with "weight > 0.4" in the TextRank algorithm and we use title effect by the ratio of one fourth. To evaluate this algorithm, we used 22 documentaries for learning purpose given in Table 14 in APPENDIX C and we categorized 18 documentaries given in Table 15 in APPENDIX C. We achieved an accuracy of 77%. In the systems which use learning mechanism, the performance of the system increases if the dataset used for learning is enlarged. Also in our

implementation, if we could use more documentaries for learning, we could get better results.

CHAPTER 4

VIDEO SUMMARIZATION

4.1 Video Summarization Algorithm

Video summarization algorithms present users a condensed version of a video. In literature, this is performed by using the image features, audio features or text features of video.

In our thesis, we wanted to use the text features of video to make summarization. There are several reasons behind this. First of all, there is a good amount of research about text summarization in natural language processing community and this research can be exploited in video summarization. Beside this, text features of video carry more semantics about the content of video than visual/audio features. Also processing of text features is more lightweight than processing of visual/audio features.

As text feature, we use the subtitles of documentary videos. We propose a video summarization approach by using text summarization techniques. Text summarization techniques constitute a summary of text by identifying the significant parts of text. We find the summary sentences of subtitle file by using some text summarization techniques. Then we find the video segments corresponding to these summary sentences. By combining the video segments of summary sentences, we create a video summary. The overall approach for video summarization is shown in Figure 14.

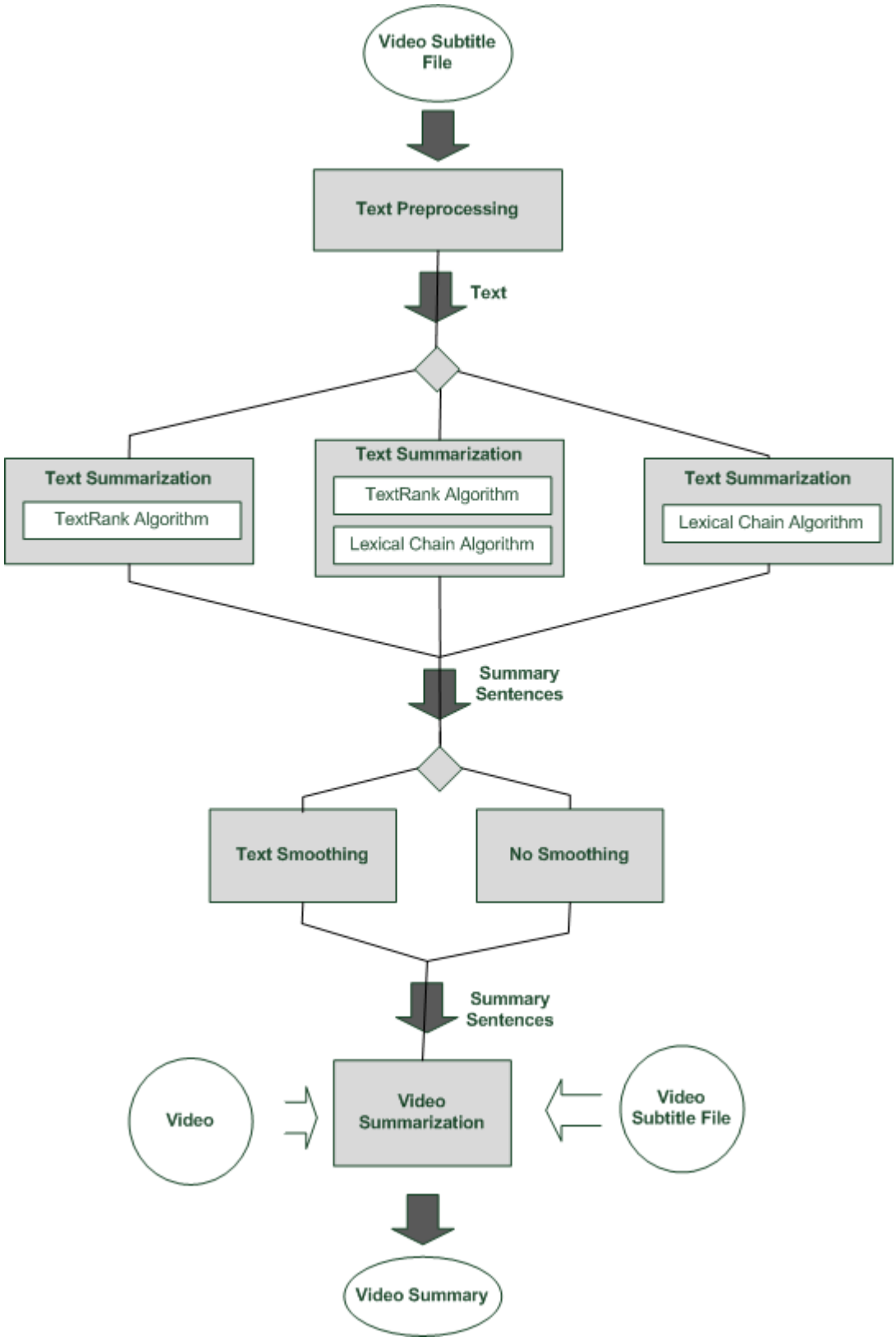

Figure 14 Overall Approach for Video Summarization

The algorithm starts with “Text Preprocessing”. Subtitle files normally contain text parts, the number of text parts and time of text parts. In this step, the text in the subtitle file is extracted and given to “Text Summarization” module. This module finds the summary sentences of the given text. There are three algorithms for finding the summary sentences; TextRank algorithm [39], Lexical Chain algorithm [62] and a combination of these two algorithms. After the summary sentences are found by one of these approaches, the output can be given to the “Text Smoothing” module. This module applies some techniques to make summary sentences more understandable and smoother. “Video Summarization” module creates video summary by using the summary sentences. The module finds the start and end times of sentences from the video subtitle file. Then the video segments corresponding to start and end times are extracted. By combining the extracted video segments, a video summary is generated. The details of these steps are given in the following sections of this chapter.

4.1.1 Text Preprocessing

In the text preprocessing step, the subtitle file is processed and the text of the subtitle file is extracted. Normally a subtitle file contains text parts, number of text parts and start time & end time of text parts. A sample subtitle file is given in Figure 15 .

```
1
00:00:25,600 --> 00:00:31,080
Human beings venture into the highest parts of our planet at their peril.
2
00:00:31,640 --> 00:00:34,480
Some might think that by climbing a great mountain
3
00:00:34,560 --> 00:00:36,320
they have somehow conquered it,
4
00:00:36,720 --> 00:00:39,800
but we can only be visitors here.
```

Figure 15 Structure of a subtitle file

As a result of this step, a text document is generated and this text document is given to text summarization module. Figure 16 shows the text generated from the subtitle file shown in Figure 15.

```
Human beings venture into the highest parts of our planet at their peril. Some might think that by climbing a great mountain they have somehow conquered it, but we can only be visitors here.
```

Figure 16 Text generated from the subtitle file in Figure 15

4.1.2 Text Summarization by TextRank Algorithm

The TextRank algorithm [39] extracts sentences for automatic summarization by identifying sentences that are more representative for the given text. To apply TextRank, we first build a graph and a vertex is added to the graph for each sentence in the text. To determine the connection between vertices, we define a "similarity" relation between them, where "similarity" is measured as a function of their content overlap. This relation can be thought as

"recommendation": a sentence mentioning about certain concepts "recommends" other sentences in the text that mention about the same concepts. Therefore a connection is made between such sentences that share common content.

The content overlap of two sentences is computed by the number of common tokens between them. To avoid promoting long sentences, the content overlap is divided by the length of each sentence.

Let two sentences S_i and S_j are given and a sentence composed of "n" words is represented by $S_i = w_1, w_2, w_3, \ldots w_n$. Then the similarity of these sentences is defined formally as;

$$Similarity(S_i, S_j) = \frac{|\{w_k \mid w_k \in S_i \,\&\, w_k \in S_j\}|}{\log(|S_i|) + \log(|S_j|)} \qquad (4.1)$$

The resulting graph is a weighted graph since the edges have a similarity weight. Then for deciding the importance of a vertex, a weighted graph-based ranking algorithm is used. Formally, the weighted score of a vertex V_i is defined as [39]:

$$WS(V_i) = (1-d) + d * \sum_{v_j \in In(v_i)} \frac{w_{ji}}{\sum_{v_k \in Out(V_j)} w_{jk}} WS(V_j) \qquad (4.2)$$

Here, $In(V_i)$ is the set of vertices that point to it and $Out(V_i)$ is the set of vertices that Vertex V_i points to. w_{ij} is the weight of the edge between the vertices V_i and V_j. d is damping factor that can be set between 0 and 1. d is usually set to 0.85 and we also use this value in our implementation.

After the ranking algorithm, sentences are sorted by using their score and top ranked sentences are selected as the summary sentences.

4.1.3 Text Summarization by Lexical Chain Algorithm

Ercan and Cicekli [63] make automated text summarization by identifying the significant sentences of text. The lexical cohesion structure of the text is exploited to determine the importance of sentences. Lexical chains can be used to analyze the lexical cohesion structure in the text.

In the proposed algorithm, first the lexical chains in the text are constructed. The lexical chaining algorithm is an implementation of the Galley et al.'s algorithm [80] with some small changes.

Then topics are roughly detected from lexical chains and the text is segmented with respect to the topics. It is assumed that the first sentence of a segment is a general description of the topic, so the first sentence of the segment is selected as the summary sentence.

4.1.4 Text Summarization by Combination of Algorithms

We propose a new summarization approach by combining the two summarization algorithms; TextRank algorithm [39] and Lexical Chain algorithm [63].

In this approach, we find the summary sentences of a text by using both the TextRank algorithm and the Lexical Chain algorithm. Afterwards, we determine the common sentences of two summaries and select these sentences to be included in the summary.

TextRank algorithm determines the summary sentences of a text in a sorted manner, that is, the summary sentences are sorted with respect to their importance score. Similarly, the Lexical Chain algorithm gives the sorted summary sentences.

After selecting the common sentences, we select the most important sentences of two algorithms up to the length of the desired summary. "Important sentences" mean "the sentences with higher importance scores". The overview of the summarization is given in Figure 17.

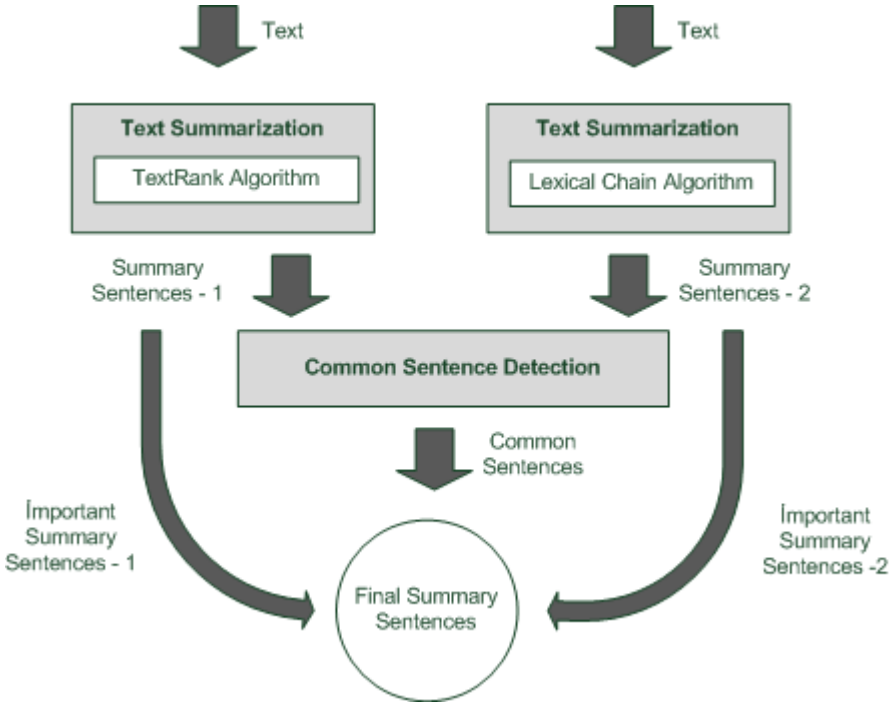

Figure 17 Overview of the text summarization by combination of algorithms

4.1.5 Text Smoothing

After text summarization, we want to make some smoothing operations to improve the understandability and completeness of the summary. Text summarization algorithms used in our summarization system selects the significant sentences of a text. It is observed that some of the selected sentences start with a pronoun and if we do not have the previous sentences in the summary, these pronouns may become confusing.

In order to handle this problem, if a sentence starts with a pronoun, the previous sentence is also included in the summary. In case that the previous sentence also starts with a pronoun, the previous sentence of that sentence is also added to the summary sentence list. When we are going back by looking at the sentences starting with a pronoun, we at most go two levels and select two more sentences. If we continue more, the length of the summary can be too long.

We observed that in case that a sentence starts with a pronoun, including the previous sentence solves the problem in most cases and the summary becomes more understandable.

4.1.6 Video Summarization

Our video summarization approach is based on the summary sentences found by text summarization algorithms. After finding the summary sentences, the start and end times of these sentences are found from video subtitle file.

For each summary sentence, the video segment corresponding to the sentence is extracted from the video by using the start and end time of the sentence. Then, by combining the extracted video parts, a video summary is created.

As evaluation domain, we select the documentary videos. In documentary videos, the speech is mostly consists of monolog and it mentions the things seen on the screen. Therefore, when we select the video segments of summary sentences, these video segments shows the objects and concepts mentioned in the sentences.

Below we give the screenshot of our video summarization system.

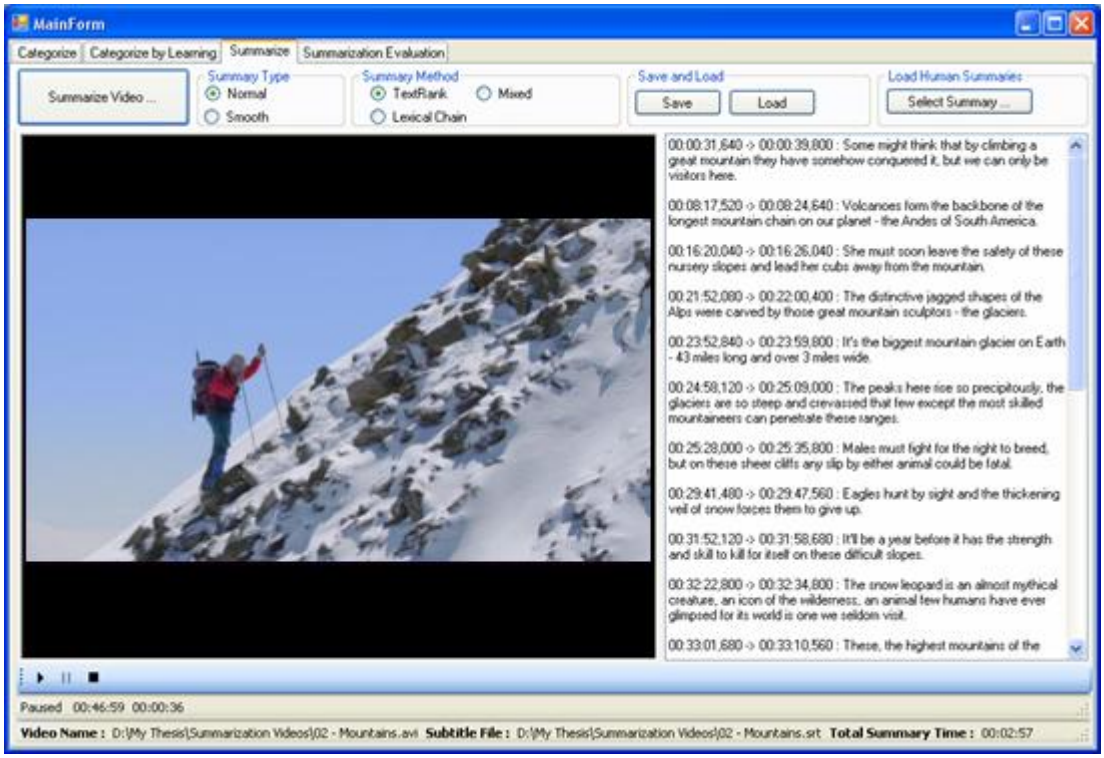

Figure 18 Video Summarization System Screenshot

In this screen, we select the summarization algorithm from the "*Summary Method*" group box; we can select "*TextRank*", "*LexicalChain*" or "*Mixed*" algorithms. From "*Summary Type*" group box, we can select "*Normal*" or "*Smooth*". If we select "*Normal*", the result of the selected summary algorithm is used directly. If we select "*Smooth*", the result of the selected summary algorithm is smoothed as discussed in Section 4.1.5. After we select summary type and summary method, we can summarize a video by using "*Summarize Video…*" button. When we click this button, we are asked to select the video file and subtitle file. Then the summary sentences and the summary video is constructed and shown on the screen. We can play, pause and stop the summary video by using the buttons below the screen.

We can "*Save*" and "*Load*" the summaries by using the "*Save and Load*" group box. Also if we have summary sentences created by humans, we can watch the summary video of humans by using the "*Load Human Summaries*" group box.

At the bottom of the screen, we can see the current "*Video Name*", "*Subtitle File*" and "*Total Summary Time*" information.

4.2 Experiments and Evaluation

Evaluation of video summaries is a hard job because summaries are subjective. Different people will compose different summaries for the same video. The evaluation of video summaries could be conducted by requesting people watch the summary and asking them several questions about the video. However, in our summarization system, since we use text summarization algorithms, we prefer to evaluate these text summarization algorithms only. We believe that success of the text summarization directly determines the success of video summarization in our system.

For the evaluation of text summarization, we use ROUGE algorithm [87] which makes evaluation by comparing the system generated output summaries to model summaries written by humans.

ROUGE (Recall-Oriented Understudy for Gisting Evaluation) [87] is the most popular summarization evaluation methodology. In all of the ROUGE metrics, it is aimed to find the percentage of overlap between the system output and the model summaries. ROUGE calculates ROUGE-N score, ROUGE-L score and ROUGE-W score. ROUGE-N score is the percentage of overlap calculated using N-grams. ROUGE-L score is calculated using LCS

(Longest Common Subsequences) and ROUGE-W score is calculated using Weighted Longest Common Subsequences.

In our video summarization system, we have used six algorithms for finding the summary of the subtitle text of a video. These algorithms can be listed as follows;

- TextRank Algorithm
- TextRank Algorithm and Smoothing the Result
- LexicalChain Algorithm
- LexicalChain Algorithm and Smoothing the Result
- A Mix of TextRank and LexicalChain Algorithms
- A Mix of TextRank and LexicalChain Algorithms and Smoothing the Result

We tried these six algorithms by using five documentaries from BBC and these documentaries are given in Appendix E. For five documentaries, we asked humans to compose summaries by selecting the most important twenty sentences from the subtitles of the documentaries. Our video categorization system also generates summaries composed of twenty sentences by using the algorithms mentioned above. We calculated ROUGE scores in order to compare the system output with human summaries, while calculating ROUGE scores; we applied Porter's Stemmer [92] on the input. We also applied a stop word list on the input. ROUGE scores of the algorithms in our video summarization system can be seen in Table 10 and Figure 19.

Table 10 ROUGE Scores of Algorithms in Our Video Summarization System

	ROUGE-1	ROUGE-2	ROUGE-3	ROUGE-4	ROUGE-L	ROUGE-W
TextRank	0,33877	0,17518	0,15327	0,13572	0,33608	0,13512
TextRank_Smooth	0,34453	0,17518	0,15327	0,13572	0,34184	0,13686
LexicalChain	0,24835	0,12915	0,10283	0,08639	0,24600	0,10413
LexicalChain_Smooth	0,25211	0,12915	0,10283	0,08639	0,24976	0,10529
Mix	0,34375	0,18500	0,15488	0,13691	0,34140	0,13934
Mix_Smooth	0,34950	0,18500	0,15488	0,13691	0,34716	0,14108

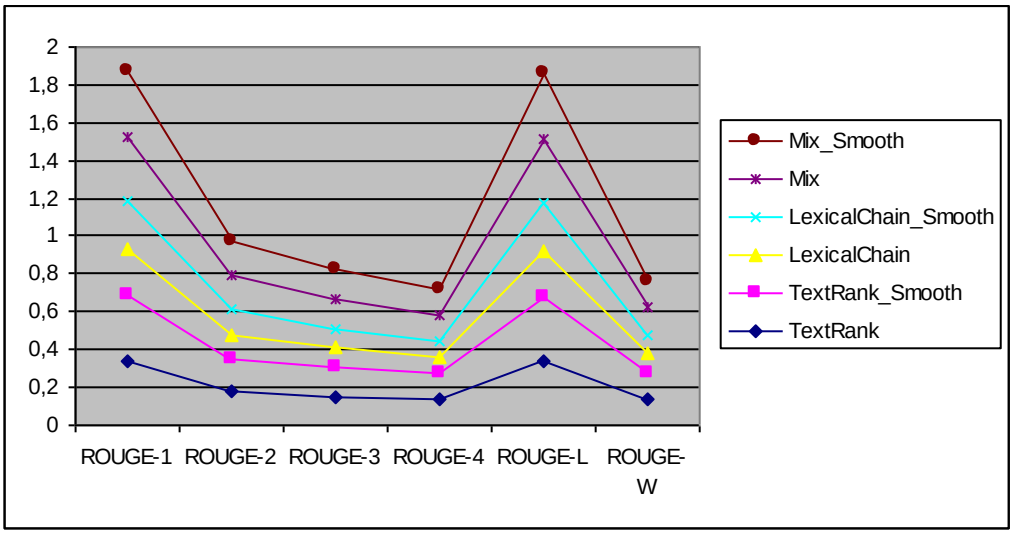

Figure 19 ROUGE Scores of Algorithms in Our Video Summarization System

We observe from Figure 19 that smoothing improves the performance of all the algorithms. When we use LexicalChain algorithm, we get better results than TextRank algorithm and we get the best results by using both TextRank and LexicalChain algorithms to find the summary sentences and smoothing the results.

CHAPTER 5

CONCLUSION

We have attacked automatic video categorization and summarization problems in this thesis work. We want to handle these problems together because their outputs support each other. Presenting both the category and the semantic summary of a video would give viewers quick and satisfactory information about that video.

We performed automatic video categorization by two video categorization methods. The first categorization method, Category Label Assignment, makes categorization by analyzing the subtitles of videos. The subtitles are processed by natural language processing techniques and WordNet lexical database and WordNet domains are used. The classification accuracy of the method is evaluated on documentary videos and promising results are obtained.

The second categorization method, Categorization by Learning, makes categorization by enabling a learning mechanism. For learning purpose, videos with known categories are utilized. However, the number of videos used for learning was limited in our system. As a future work, we want to improve the learning by using more videos. It is known that using more data for learning increases the performance of the system and gives better results.

We perform video summarization by using video subtitles and employing text summarization methods. We use two well-known text summarization algorithms (Mihalcea and Tarau [39], Ercan and Cicekli [63]) and apply the results to video summarization domain. In this work, we take the advantage of the characteristics of the documentary videos. In documentary videos, the speech and the display of the video have a strong correlation in the way that mostly both of them give information about the same entities.

Video summary is produced by extracting the video parts corresponding to the summary sentences. Video parts extraction could be improved by employing a shot identification mechanism. An extracted video part could be extended by finding the start and end of the residing shot. By this way, the video parts could show a more complete presentation.

In video summarization evaluation, we evaluate the text summaries of videos. We compare the program summaries with human generated summaries and find the ROUGE score of program summaries. As a future work, we want to perform the evaluation by using the video summaries not only text summaries. Video summaries could be watched by viewers and the viewers could evaluate the results.

Both of our algorithms are currently for English, but it is possible to convert these algorithms to different languages. The language dependency of the algorithms is caused by the WordNet and the NLP tools such as pos tagger. When the WordNet and the required NLP tools are available for Turkish, our video categorization and summarization algorithms can be used for videos with Turkish subtitles.

REFERENCES

1. Brezeale, D., Cook, D.J.: Automatic Video Classification: A Survey of the Literature. IEEE Trans. Systems, Man, and Cybernetics-Part C: Applications and Reviews 38(3), 416–430 (2008) 16. Lu, L., Zhang, H.

2. Polyxeni Katsiouli, Vassileios Tsetsos, Stathes Hadjiefthymiades. Semantic Video Classification Based on Subtitles Domain Terminologies. University of Athens, Panepistimioupolis, Ilissia 15784, Greece.

3. W.Zhu, C.Toklu, and S-P.Liou, "Automatic News Video Segmentation and Categorization Based on Closed-Captioned Text", ISIS technical report series, Vol 2001-20, Dec. 2001.

4. D. Brezeale and D. J. Cook "Using closed captions and visual features to classify movies by genre," in presented at the, *Poster Session 7th Int. Workshop Multimedia Data Min. (MDM/KDD)* San Jose, CA, 2006.

5. Bentivogli, L., Forner, P., Magnini, B., Pianta, E.: Revising WordNet Domains Hierarchy: Semantics, Coverage, and Balancing. In Proceedings of COLING Workshop on Multilingual Linguistic Resources, Geneva Switzerland (2004) 101-108

6. K. P. Bennett and C. Campbell. Support vector machines: Hype or hallelujah? SIGKDD Explorations, 2(2):1–13, 2000.

7. W.-H. Lin and A. Hauptmann, "News video classification using SVMbased multimodal classifiers and combination strategies," in *Proc. ACM Multimedia*, 2002, pp. 323–326.

8. P. Wang, R. Cai, and S.-Q. Yang, "A hybrid approach to news video classification multimodal features," in Proc. Joint Conf. 4th Int. Conf. Inf., Commun. Signal Process. 4th Pacific Rim Conf. Multimedia, 2003, pp. 787–791.

9. X. Yuan, W. Lai, T. Mei, X.-S. Hua, X.-Q. Wu, and S. Li, "Automatic video genre categorization using hierarchical SVM," in Proc. IEEE Int. Conf. Image Process. (ICIP), 2006, pp. 2905–2908.

10. G.Wei, L. Agnihotri, and N. Dimitrova, "TV program classification based on face and text processing," in Proc. IEEE Int. Conf. Multimedia Expo., 2000, vol. 3, pp. 1345–1348.

11. N. Dimitrova, L. Agnihotri, and G. Wei, "Video classification based on HMM using text and faces," presented at the Eur. Signal Process. Conf. (EUSIPCO 2000), Tampere, Finland, Sep. 2000.

12. R. S. Jasinschi and J. Louie, "Automatic TV program genre classification based on audio patterns," in Proc. IEEE 27th Euromicro Conf., 2001, pp. 370–375.

13. M. H. Lee, S. Nepal, and U. Srinivasan, "Edge-based semantic classification of sports video sequences," in Proc. Int. Conf. Multimedia Expo. (ICME 2003), vol. 2, pp. 157–160.

14. V. Kobla, D. DeMenthon, and D. Doermann, "Identifying sports videos using replay, text, and camera motion features," in Proc. SPIE Conf. Storage Retrieval Media Databases, 2000, pp. 332–343.

15. X. Gibert, H. Li, and D. Doermann, "Sports video classification using HMMs," in Proc. Int. Conf. Multimedia Expo (ICME 2003), vol. 2, pp. 345–348.

16. J. Huang, Z. Liu, Y. Wang, Y. Chen, and E. K. Wong, "Integration of multimodal features for video scene classification based on HMM," in Proc. 3rd IEEE Workshop Multimedia Signal Process., 1999, pp. 53–58.

17. L.-Q. Xu and Y. Li, "Video classification using spatial-temporal features and PCA," in Proc. Int. Conf. Multimedia Expo (ICME 2003), pp. 485–488.

18. G. Y. Hong, B. Fong, and A. Fong, "An intelligent video categorization engine," Kybernetes, vol. 34, no. 6, pp. 784–802, 2005.

19. Z. Rasheed, Y. Sheikh, andM. Shah, "Semantic film preview classification using low-level computable features," presented at the 3rd Int. Workshop Multimedia Data Document Eng. (MDDE 2003), Berlin, Germany.

20. J.-Y. Pan and C. Faloutsos, "Videocube: A novel tool for video mining and classification," presented at the Int. Conf. Asian Digit. Libr., Singapore, 2002.

21. A. Hauptmann, R. Yan, Y. Qi, R. Jin, M. Christel, M. Derthick, M.-Y. Chen, R. Baron, W.-H. Lin, and T. D. Ng, "Video classification and retrieval with the informedia digital video library system," presented at the Text Retrieval Conf. (TREC 2002), Gaithersburg, MD.

22. Z. Rasheed andM. Shah, "Movie genre classification by exploiting audiovisual features of previews," in Proc. IEEE Int. Conf. Pattern Recognit., 2002, vol. 2, pp. 1086–1089.

23. G. Iyengar and A. Lippman, "Models for automatic classification of video sequences," in Proc. SPIE Storage Retrieval Image Video Databases VI, I. K. Sethi and R. C. Jain, Eds., 1997, vol. 3312, pp. 216–227.

24. B. T. Truong, C. Dorai, and S. Venkatesh, "Automatic genre identification for content-based video categorization," in Proc. 15th Int. Conf. Pattern Recognit., 2000, vol. IV, pp. 230–233.

25. R. Jadon, S. Chaudhury, and K. Biswas, "Generic video classification: An evolutionary learning based fuzzy theoretic approach," presented at the Indian Conf. Comput. Vis. Graph. Image Process. (ICVGIP), Ahmedabad, India, 2002.

26. S. Fischer, R. Lienhart, andW. Effelsberg, "Automatic recognition of film genres," in Proc. 3rd ACM Int. Conf. Multimedia (MULTIMEDIA 1995), pp. 295–304.

27. P. Wang, R. Cai, and S.-Q. Yang, "A hybrid approach to news video classification multimodal features," in Proc. Joint Conf. 4th Int. Conf. Inf., Commun. Signal Process. 4th Pacific Rim Conf. Multimedia, 2003, pp. 787–791.

28. M. Roach, J.Mason, andM. Pawlewski, "Video genre classification using dynamics," IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP 2001), vol. 3, pp. 1557–1560.

29. M. Roach, J. S. Mason, and M. Pawlewski, "Motion-based classification of cartoons," in Proc. Int. Symp. Intell. Multimedia, 2001, pp. 146–149.

30. M. Roach, J. Mason, and L.-Q. Xu, "Video genre verification using both acoustic and visual modes," in Proc. Int. Workshop Multimedia Signal Process., 2002, pp. 157–160.

31. E. Wold, T. Blum, D. Keislar, and J. Wheaton, "Content-based classification, search, and retrieval of audio," IEEE MultiMedia, vol. 3, no. 3, pp. 27–36, Fall 1996.

32. P. Q. Dinh, C. Dorai, and S. Venkatesh, "Video genre categorization using audio wavelet coefficients," presented at the 5th Asian Conf. Comput.Vis., Melbourne, Australia, 2002.

33. J. Fan, H. Luo, J. Xiao, and L. Wu, "Semantic video classification and feature subset selection under context and concept uncertainty," in Proc. 4th ACM/IEEE-CS Joint Conf. Digit. Libr. (JCDL 2004), pp. 192–201.

34. S. Moncrieff, S. Venkatesh, and C. Dorai, "Horror film genre typing and scene labeling via audio analysis," in Proc. Int. Conf. Multimedia Expo (ICME 2003), vol. 1, pp. 193–196.

35. Z. Liu, Y. Wang, and T. Chen, "Audio feature extraction and analysis for scene segmentation and classification," J. VLSI Signal Process. Syst., vol. 20, no. 1/2, pp. 61–79, 1998.

36. Z. Liu, J. Huang, and Y. Wang, "Classification of TV programs based on audio information using hidden Markov model," in Proc. IEEE Signal Process. Soc. Workshop Multimedia Signal Process., 1998, pp. 27–32.

37. L. He, E. Sanocki, A. Gupta, and J. Grudin, "Auto-summarization of audio-video presentations," in Proc. 7th ACM Int. Conf. Multimedia (Part 1) (MULTIMEDIA 1999), pp. 489–498.

38. W. Qi, L. Gu, H. Jiang, X.-R. Chen, and H.-J. Zhang, "Integrating visual, audio and text analysis for news video," in Proc. 7th IEEE Int. Conf. Image Process. (ICIP), Sep. -2000, pp. 520–523.

39. R. Mihalcea and P. Tarau. 2004. TextRank - bringing order into texts. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2004), Barcelona, Spain.

40. Banerjee, S., Pedersen, T.: An Adapted Lesk Algorithm for Word Sense Disambiguation Using WordNet. In the Proceedings of the 3rd International Conference on Intelligent Text Processing and Computational Linguistics (CICLING-02) Mexico City, Mexico (2002)

41. Stanford Log-linear Part-Of-Speech Tagger, http://nlp.stanford.edu/software/tagger.shtml, Last Access Date: 12 May, 2009

42. Penn Treebank, http://www.cis.upenn.edu/~treebank/, Last Access Date: 15 May, 2009

43. Lesk, M.: Automatic sense disambiguation using machine readable dictionaries: How to tell a pine cone from a ice cream cone. Proceedings of the 5th Annual International Conference on Systems Documentation (1986)

44. Brin S., Page L.: The anatomy of a large-scale hypertextual Web search engine. Computer Networks and ISDN Systems. Vol 30. No. 1-7 (1998) 107-117

45. WordNet Domains, http://wndomains.fbk.eu/, Last Access Date: 1 September, 2009

46. Tf-idf Definition, http://en.wikipedia.org/wiki/Tf%E2%80%93idf, Last Access Date: 23 August, 2009

47. Cosine Similarity Definition, http://en.wikipedia.org/wiki/Cosine_similarity, Last Access Date: 25 August, 2009

48. R. Brandow, K. Mitze, and Lisa F. Rau. Automatic condensation of electronic publications by sentence selection. Inf. Process. Manage., 31(5):675–685, 1995.

49. H. P. Edmundson. New methods in automatic extracting. J. ACM, 16(2):264–285, 1969.

50. Julian Kupiec, Jan O. Pedersen, and Francine Chen. A trainable document summarizer. In SIGIR'95, pages 68–73. ACM Press, 1995.

51. Simone Teufel and Marc Moens. Sentence extraction as a classification task. In Inderjeet Mani and Mark T. Maybury, editors, ACL/EACL97-WS, Madrid, Spain, 1997.

52. Dragomir R. Radev, Hongyan Jing, and Malgorzata Budzikowska. Centroid based summarization of multiple documents: Sentence extraction, utilitybased evaluation, and user studies. In Udo Hahn, Chin-Yew Lin, Inderjeet Mani, and Dragomir R. Radev, editors, ANLP/NAACL00-WS, Seattle, WA, April 2000.

53. G¨unes Erkan and Dragomir R. Radev. Lexrank: Graph-based lexical centrality as salience in text summarization. J. Artif. Intell. Res. (JAIR), 22:457–479, 2004.

54. Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford Digital Library Technologies Project, 1998.

55. Regina Barzilay and Michael Elhadad. Using Lexical Chains for Text Summarization. In Inderjeet Mani and Mark T. Maybury, editors, Advances in Automatic Text Summarization, pages 111–121. The MIT Press, 1999.

56. Gregory H. Silber and Kathleen McCoy. Efficient text summarization using lexical chains. In Proceedings of the ACM Conference on Intelligent User Interfaces (IUI'2000), January 9–12 2000.

57. Meru Brunn, Yllias Chali, and Christopher J. Pinchak. Text summarization using lexical chains. New Orleans, LA, 2001.

58. Yllias Chali and Maheedhar Kolla. University of lethridge summarizer at duc04. In DUC04, Boston, USA, July 2004.

59. William P. Doran, Nicola Stokes, Joe Carthy, and John Dunnion. Assessing the impact of lexical chain scoring methods and sentence extraction schemes on summarization. In CICLing, pages 627–635, 2004.

60. E. Newman J. Dunnion J. Carthy F. Toolan W. Doran, N. Stokes. News story gisting at university college dublin. In DUC04, Boston, USA, July 2004.

61. Jian-Yun Nie Quan Zhou Le Sun. Is sum: A multi-document summarizer based on document index graphic and lexical chains. In DUC05, Vancouver, CA, July 2005.

62. Tat-Seng Chua Shiren Ye, Long Qiu and Min-Yen Kan. Nus at duc 2005: Understanding documents via concept links. In DUC05, Vancouver, CA, July 2005.

63. Ercan, G. and I. Cicekli (2008). Lexical cohesion based topic modeling for summarization. In Proceedings of the Cicling 2008, pp. 582–592.

64. A. Ekin, A.M. Tekalp, R. Mehrotra, Automatic soccer video analysis and summarization, IEEE Transactions on Image Processing 12 (7) (2003) 796–807.

65. W. Cheng, D. Xu, An approach to generating two-level video abstraction, in: Proceedings of the 2nd IEEE International Conference on Machine Learning and Cybernetics, vol. 5, Xi-an, China, 2–5 November, 2003, pp. 2896–2900.

66. A. Girgensohn, A fast layout algorithm for visual video summaries, in: Proceedings of the IEEE International Conference on Multimedia and Expo (ICME '03), vol. 2, Baltimore, MD, USA, 6–9 July, 2003, pp. 77–80.

67. C. Ngo, Y. Ma, H. Zhang, Video summarization and scene detection by graph modeling, IEEE Transactions on Circuits and Systems for Video Technology 15 (2) (2005) 296–305.

68. Z. Cernekova, I. Pitas, C. Nikou, Information theory-based shot cut/ fade detection and video summarization, IEEE Transactions on Circuits and Systems for Video Technology 16 (1) (2006) 82–91.

69. Z. Li, G.M. Schuster, A.K. Katsaggelos, MINMAX optimal video summarization, IEEE Transactions on Circuits and Systems for Video Technology 15 (10) (2005) 1245–1256.

70. G. Ciocca, R. Schettini, Dynamic storyboards for video content summarization, in: Proceedings of the 8th ACM SIGMM International Workshop on Multimedia Information Retrieval, Santa Barbara, CA, 26–27 October, 2006.

71. N. Benjamas, N. Cooharojananone, C. Jaruskulchai, Flashlight and player detection in fighting sport for video summarization, in: Proceedings of the IEEE International Symposium on Communications and Information Technology (ISCIT 2005), vol. 1, Beijing, China, 12–14 October 2005, pp. 441–444.

72. B. Jung, T. Kwak, J. Song, Y. Lee, Narrative abstraction model for story-oriented video, in: Proceedings of the 12th Annual ACM International Conference on Multimedia, New York, NY, USA, 10–15 October, 2004.

73. T. Mei, C. Zhu, H. Zhou, X. Hua, Spatio-temporal quality assessment for home videos, in: Proceedings of the 13th Annual ACM International Conference on Multimedia, Singapore, 6–11 November, 2005.

74. F.N. Bezerra, E. Lima, Low cost soccer video summaries based on visual rhythm, in: Proceedings of the 14th Annual ACM International Conference on Multimedia, Santa Barbara, CA, 23–27 October, 2006, pp. 71–77.

75. M. Xu, C. Maddage, C. Xu, M. Kankanhalli, Q. Tian, Creating audio keywords for event detection in soccer video, in: Proceedings of the IEEE

International Conference on Multimedia and Expo (ICME '03), vol. 2, Baltimore, USA, 6–9 July, 2003, pp. 281–284.

76. Y. Rui, A. Gupta, A. Acero, Automatically extracting highlights for TV baseball programs, in: Proceedings of the 8th ACM International Conference on Multimedia, Los Angeles, CA, USA, 30 October, 2000, pp. 105–115.

77. A. Money and H. Agius. Video summarisation: A conceptual framework and survey of the state of the art. Journal of Visual Communication and Image Representation, 19(2):121{143, February 2008.

78. Pickering, M., Wong, L., and Ruger, S. (2003). ANSES: Summarisation of News Video. In: Proceedings of CIVR-2003, University of Illinois, IL, USA, July 24-25, 2003.

79. Tsvetomira Tsoneva , Mauro Barbieri , Hans Weda, Automated summarization of narrative video on a semantic level, Proceedings of the International Conference on Semantic Computing, p.169-176, September 17-19, 2007

80. Galley, M.,McKeown, K.: Improving word sense disambiguation in lexical chaining. In: Proceedings of the 18th International Joint Conference on Artificial Intelligence (IJCAI 2003), pp. 1486–1488 (2003)

81. K. Aas, and L. Eikvil: Text Categorisation: A Survey. Norwegian Computing Center, Oslo, 1999

82. Y. Yang, and J. O. Pedersen. A Comparative Study on Feature Selection in Text Categorization. In Proceedings of the Fourteenth International

Conference on Machine Learning.. D. H. Fisher, Ed. Morgan Kaufmann Publishers, San Francisco, CA (1997) 412-420

83. Yang, Y.: Expert network: Effective and Efficient learning from human decisions in text categorization and retrieval. In 17th Ann Int. ACM SIGIR Conference on Research and Development in Information Retrieval (1994) 13-22

84. Yang, Y., Chute, C.G.: An example-based mapping method for text categorization and retrieval. ACM Transaction on Information Systems (TOIS) (1994) 253-277

85. Kwok, J. T.-Y.: Automated Text Categorization Using Support Vector Machine. In Proceedings of the International Conference on Neural Information Processing. Kitakyushu Japan (1998) 347-351

86. 16. Hepple, M.: Independence and Commitment: Assumptions for Rapid Training and Execution of Rule-based Part-of-Speech Taggers. Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics. Hong Kong (2000)

87. Chin-Yew Lin and Eduard H. Hovy. Automatic evaluation of summaries using n-gram co-occurrence statistics. In HLT-NAACL, 2003.

88. S. W. Smoliar and H. J. Zhang. Content-based video indexing and retrieval. IEEE Multimedia, pages 62–72, 1994.

89. D. DeMenthon, V. Kobla, and D. Doermann. Video summarization by curve simplification. Proceedings of ACM Multimedia, pages 211–218, 1998.

90. B. Barbieri, N. Dimitrova, and L. Agnihotri. Movie-in-aminute: Automatically generated video previews. IEEE Pacific Rim Conference on Multimedia, 2:9–18, 2004.

91. K. Fujimura, K. Honda, and K. Uehara. Automatic video summarization by using color and utterance information. Proc. of IEEE ICME, pages 49–52, 2002.

92. Porter, M.F. 1980. An Algorithm for Suffix Stripping. Program, 14, pp. 130-137.

APPENDIX A

POS TAG ABBREVIATIONS AND PENN TREEBANK DESCRIPTIONS

Table 11 Pos Tag Abbreviations and Penn Treebank Descriptions

POS Tag Abbreviation	Penn Treebank Description
CC	Coordinating conjunction
CD	Cardinal number
DT	Determiner
EX	Existential there
FW	Foreign word
IN	Preposition/subordinate conjunction
JJ	Adjective
JJR	Adjective, comparative
JJS	Adjective, superlative
LS	List item marker
MD	Modal
NN	Noun, singular or mass
NNP	Proper noun, singular
NNPS	Proper noun, plural
NNS	Noun, plural
PDT	Predeterminer
POS	Possessive ending
PRP	Personal pronoun
PRP$	Possessive pronoun
RB	Adverb
RBR	Adverb, comparative
RBS	Adverb, superlative
RP	Particle
SYM	Symbol
TO	to
UH	Interjection
VB	Verb, base form
VBD	Verb, past tense
VBG	Verb, gerund/present participle
VBN	Verb, past participle

Table 11 Pos Tag Abbreviations and Penn Treebank Descriptions (Continued)

VBP	Verb, non-3rd ps. sing. Present
VBZ	Verb, 3rd ps. sing. Present
WDT	wh-determiner
WP	wh-pronoun
WP$	Possessive wh-pronoun
WRB	wh-adverb
``	Left open double quote
,	Comma
"	Right close double quote
.	Sentence-final punctuation
:	Colon, semi-colon
$	Dollar sign
#	Pound sign
-LRB-	Left parenthesis
-RRB-	Right parenthesis

APPENDIX B

STOP WORDS LIST

Table 12 Stop Words List

a	but	further	mostly	several	towards
about	by	get	move	she	twelve
above	call	give	much	should	twenty
across	can	go	must	show	two
after	cannot	had	my	side	un
afterwards	cant	has	myself	since	under
again	co	hasnt	name	sincere	until
against	computer	have	namely	six	up
all	con	he	neither	sixty	upon
almost	could	hence	never	so	us
alone	couldnt	her	nevertheless	some	very
along	cry	here	next	somehow	via
already	de	hereafter	nine	someone	was
also	describe	hereby	no	something	we
although	detail	herein	nobody	sometime	well
always	do	hereupon	none	sometimes	were
am	done	hers	noone	somewhere	what
among	down	herself	nor	still	whatever
amongst	due	him	not	such	when
amoungst	during	himself	nothing	system	whence
amount	each	his	now	take	whenever
an	eg	how	nowhere	ten	where
and	eight	however	of	than	whereafter
another	either	hundred	off	that	whereas
any	eleven	i	often	the	whereby
anyhow	else	ie	on	their	wherein
anyone	elsewhere	if	once	them	whereupon
anything	empty	in	one	themselves	wherever
anyway	enough	inc	only	then	whether
anywhere	etc	indeed	onto	thence	which
are	even	interest	or	there	while

Table 12 Stop Words List (Continued)

around	ever	into	other	thereafter	Whither
as	every	is	others	thereby	who
at	everyone	it	otherwise	therefore	whoever
back	everything	its	our	therein	whole
be	everywhere	itself	ours	thereupon	whom
became	except	keep	ourselves	these	whose
because	few	last	out	they	why
become	fifteen	latter	over	thick	will
becomes	fify	latterly	own	thin	with
becoming	fill	least	part	third	within
been	find	less	per	this	without
before	fire	ltd	perhaps	those	would
beforehand	first	made	please	though	yet
behind	five	many	put	three	you
being	for	may	rather	through	your
below	former	me	re	throughout	yours
beside	formerly	meanwhile	same	thru	yourself
besides	forty	might	see	thus	yourselves
between	found	mill	seem	to	
beyond	four	mine	seemed	together	
bill	from	more	seeming	too	
both	front	moreover	seems	top	
bottom	full	most	serious	toward	

APPENDIX C

DOCUMENTARIES USED FOR VIDEO CATEGORIZATION EVALUATION

Table 13 Documentaries Used in Evaluation

Documentary Type	Documentary Name	Expert Category
BBC	Wildlife Specials - Leopard	Animals
BBC	Wildlife Specials - Serpent	Animals
BBC	Wildlife Specials - Tiger	Animals
National Geographic	Those Wonderful Dogs	Animals
National Geographic	The New Chimpanzees	Animals
National Geographic	The Secret Life of Cats	Animals
BBC	War of the Century - 01	War
BBC	War of the Century - 02	War
BBC	War of the Century - 03	War
BBC	War of the Century - 04	War
National Geographic	The Battle for Midway	War
National Geographic	Untold Stories of WW II	War
BBC	Pearl Harbor Legacy of Attack - 01.srt	War
BBC	Pearl Harbor Legacy of Attack - 02.srt	War
BBC	Planet Earth - From Pole to Pole	Geography
BBC	Planet Earth - Mountains	Geography
BBC	Planet Earth - Fresh Water	Geography
BBC	Planet Earth - Caves	Geography
BBC	Planet Earth - Deserts	Geography
National Geographic	Code of the Maya Kings	Geography
National Geographic	Lost Ships of the Mediterrian	Geography
National Geographic	The Everest	Geography
National Geographic	The Silk Road	Geography
BBC	God on the Brain	Religion
PBS	Islam : Empire of Faith - 01	Religion

Table 13 Documentaries Used in Evaluation (Continued)

PBS	Islam : Empire of Faith - 02	Religion
BBC	The Life of Buddha	Religion
BBC	Samba to Bossa	Music
BBC	The Pink Floyd Story	Music
BBC	Tropicalia Revolution	Music
Other	The Power of Nightmares - Baby Its Cold Outside	Politics
Other	The Power of Nightmares - The Phantom Victory	Politics
Other	The Power of Nightmares - The Shadows in the Cave	Politics
National Geographic	The Incredible Human Body	Science
Other	The Struggle Against Cancer	Science
National Geographic	Mysteries of Egypt	History
BBC	The Art of Spain - The Moorish South	Art
BBC	The Art of Spain - The Dark Heart	Art
BBC	The Art of Spain - The Mystical North	Art
Other	Charles Lindbergh - The Lone Eagle	People

Table 14 Documentaries Used for Learning

Documentary Type	**Documentary Name**	**Expert Category**
BBC	Wildlife Specials - Leopard	Animals
National Geographic	Those Wonderful Dogs	Animals
National Geographic	The Secret Life of Cats	Animals
BBC	War of the Century - 01	War
BBC	War of the Century - 02	War
National Geographic	The Battle for Midway	War
BBC	Pearl Harbor Legacy of Attack - 01.srt	War
BBC	Planet Earth - From Pole to Pole	Geography
BBC	Planet Earth - Mountains	Geography
National Geographic	Code of the Maya Kings	Geography
National Geographic	Lost Ships of the Mediterrian	Geography
PBS	Islam : Empire of Faith - 01	Religion
PBS	Islam : Empire of Faith - 02	Religion
BBC	The Pink Floyd Story	Music
BBC	Tropicalia Revolution	Music

Table 14 Documentaries Used for Learning (Continued)

Other	The Power of Nightmares - The Phantom Victory	Politics
Other	The Power of Nightmares - The Shadows in the Cave	Politics
National Geographic	The Incredible Human Body	Science
National Geographic	Mysteries of Egypt	History
BBC	The Art of Spain - The Moorish South	Art
BBC	The Art of Spain - The Dark Heart	Art
Other	Charles Lindbergh - The Lone Eagle	People

Table 15 Documentaries Used for Learning Evaluation

Documentary Type	**Documentary Name**	**Expert Category**
BBC	Wildlife Specials - Serpent	Animals
BBC	Wildlife Specials - Tiger	Animals
National Geographic	The New Chimpanzees	Animals
BBC	War of the Century - 03	War
BBC	War of the Century - 04	War
National Geographic	Untold Stories of WW II	War
BBC	Pearl Harbor Legacy of Attack - 02.srt	War
BBC	Planet Earth - Fresh Water	Geography
BBC	Planet Earth - Caves	Geography
BBC	Planet Earth - Deserts	Geography
National Geographic	The Everest	Geography
National Geographic	The Silk Road	Geography
BBC	God on the Brain	Religion
BBC	The Life of Buddha	Religion
BBC	Samba to Bossa	Music
Other	The Power of Nightmares - Baby Its Cold Outside	Politics
Other	The Struggle Against Cancer	Science
BBC	The Art of Spain - The Mystical North	Art

APPENDIX D

MATRIX REPRESENTING THE DOMAIN DISTRIBUTION OF CATEGORIES

Table 16 Matrix Representing the Domain Distribution of Categories

	Geography	Animals	Politics	History	Religion	War
geography	0,0552444	0,032574	0,039201	0,042323	0,050815	0,050245
animals	0,0302953	0,053447	0,009224	0,020669	0,009907	0,017157
biology	0,0315682	0,041429	0,016141	0,024606	0,016299	0,019914
entomology	0,0022912	0,000949	0	0,001969	0,000639	0,001838
politics	0,0068737	0,008223	0,037663	0,005906	0,012784	0,016238
psychology	0,0043279	0,007906	0,008455	0,00689	0,007031	0,005208
history	0,0099287	0,008223	0,01691	0,014764	0,020773	0,020833
time_period	0,0313136	0,023087	0,017294	0,025591	0,020134	0,029105
religion	0,0089104	0,004744	0,021522	0,014764	0,040588	0,006127
transport	0,0129837	0,012334	0,013451	0,008858	0,008949	0,024816
commerce	0,0022912	0,004428	0,004612	0,000984	0,00767	0,001532
enterprise	0,0020367	0,001581	0,003843	0,003937	0,004794	0,003676
nautical	0,0071283	0,004111	0,003459	0,009843	0,003516	0,017157
sport	0,0040733	0,007906	0,003075	0,004921	0,003835	0,006127
play	0,0048371	0,002214	0,000769	0,004921	0,001278	0,002757
swimming	0,0002546	0,000316	0,000384	0	0	0,001532
military	0,0132383	0,013283	0,026134	0,011811	0,024609	0,060662
medicine	0,0066191	0,010436	0,004612	0,00689	0,00767	0,007047
mathematics	0,0022912	0,001265	0,00269	0,002953	0,001918	0,001532
music	0,0071283	0,003163	0,004228	0,002953	0,005113	0,005515
linguistics	0,0053462	0,002214	0,004996	0,001969	0,007031	0,004902
literature	0,0106925	0,00759	0,006149	0,007874	0,007031	0,004902
art	0,0033096	0,003795	0,001537	0,005906	0,006072	0,001838
painting	0,0020367	0,000316	0	0,001969	0,000959	0,000919
graphic_arts	0	0	0	0	0	0
engineering	0,0012729	0,000949	0,000769	0,000984	0,000639	0,000306
industry	0,0063646	0,00253	0,003075	0,009843	0,007351	0,004596
computer_science	0,0040733	0,002214	0,001922	0,000984	0,001598	0,003983
sociology	0,009165	0,006958	0,013451	0,005906	0,011825	0,008578
person	0,0239308	0,034472	0,031514	0,029528	0,030681	0,034007

Table 16 Matrix Representing the Domain Distribution of Categories (Continued)

	Science	Music	Art	People
geography	0,019861	0,03338	0,038857	0,049541
animals	0,016882	0,018544	0,010286	0,017431
biology	0,020854	0,025962	0,016381	0,021101
entomology	0,000993	0	0,001143	0
politics	0,007944	0,018081	0,011048	0,005505
psychology	0,006951	0,006027	0,005714	0,006422
history	0,00993	0,011127	0,021333	0,013761
time_period	0,02284	0,025498	0,017524	0,027523
religion	0,010924	0,011127	0,049143	0,009174
transport	0,017875	0,012054	0,009905	0,027523
commerce	0,000993	0,005563	0,003048	0,006422
enterprise	0,000993	0,003245	0,001905	0,002752
nautical	0,002979	0,001391	0,002286	0,007339
sport	0,013903	0,006954	0,001905	0,005505
play	0,00993	0,0051	0,001143	0,00367
swimming	0	0	0	0
military	0,011917	0,013908	0,020571	0,016514
medicine	0,021847	0,0051	0,005333	0,006422
mathematics	0,008937	0,002318	0,003048	0,004587
music	0,008937	0,037552	0,007619	0,00367
linguistics	0,005958	0,012981	0,006095	0,001835
literature	0,007944	0,012054	0,008762	0,007339
art	0,004965	0,009272	0,010286	0,001835
painting	0	0,000464	0,005333	0
graphic_arts	0	0	0	0
engineering	0,000993	0	0	0,000917
industry	0,002979	0,003709	0,004571	0,008257
computer_science	0,007944	0,0051	0,001524	0,001835
sociology	0,006951	0,015299	0,01181	0,007339
person	0,043694	0,030598	0,033524	0,033028

APPENDIX E

DOCUMENTARIES USED FOR VIDEO SUMMARIZATION EVALUATION

Table 17 Documentaries Used for Video Summarization Evaluation

Documentary Type	Documentary Name	Expert Category
BBC	Planet Earth - From Pole to Pole	Geography
BBC	Planet Earth - Fresh Water	Geography
BBC	Planet Earth - Deserts	Geography
BBC	Wildlife Specials - Leopard	Animals
BBC	Wildlife Specials - Serpent	Animals