Republic of Iraq
Ministry of Higher Education and Scientific Research
University of Diyala
Department of Computer Science

Car Surveillance Video Summarization Model Using Car Plate Detection

A Research

Submitted to the Department of Computer Science\ College of Sciences\ University of Diyala in a Partial Fulfillment of the Requirements for the Degree of Master in Computer Science.

By

Nouria Kareem Khursheed

Supervised By

Prof. Dr. Ziyad Tariq Mustafa Al-Ta'i

2021 **A.C** 1442 **A.H**

بَسم الله الرحمن الرحيم

اقْرَأْ بِاسْمِ رَبِّكَ الَّذِي خَلَقَ (1) خَلَقَ الْإِنسَانَ مِنْ عَلَقٍ

(2) اقْرَأْ وَرَبُّكَ الْأَكْرَمُ (3) الَّذِي عَلَّمَ بالْقَلَمِ (4) عَلَّمَ

الْإِنسَانَ مَا لَمْ يَعْلَمْ (5)

صدق الله العظيم

سُورَةُ العَلَق

ACKNOWLEDGMENT

First of all, praise is to Allah Who helped me and gave me the ability to achieve this research from first to last steps.

I would like to express my deep gratitude and sincere thanks to my supervisor **Prof. Dr. Ziyad Tariq Mustafa** *for his guidance, assistance, and encouragement during the course of this project.*

Great thanks are due to the lecturers in the Department of Computer Science for the continuous support during the period of my studies.

Deep gratitude and special thanks are extended to my family for their encouragement and support for me to succeed in doing this work.

Special thanks to all my friends for giving me advice.

Nouria

Dedication

To my family with my love

Abstract

Today, video is a common medium for sharing information. Navigating the internet to download a certain form of video, takes a long time, a lot of bandwidth, and a lot of disk space. Since sending video over the internet is too costly, therefore video summarization has become a critical technology.

Monitoring vehicles of people from a security and traffic perspective is a major issue. This monitoring depends on the identification of the license plate of vehicles.

In this thesis, the proposed system includes two parts: first, a video summary that contains all the cars shown in the video, and the second is to define the license plate and summarize the video. The First part, contains training and testing stages. Video summarization training comprises video preprocessing, Viola-Jones training with False Alarm Rate and Number of Cascade stage, for optimization Support Vector Machine (SVM) with Local Binary Pattern (LBP) features extraction with outlier and kernel scale parameters. Video summarization testing contains: test video preprocessing, car plate (detection, cropping, resizing, and grouping), and viewing related frames. The second part which is used to define the car plate to summarize the video contains training and testing stages. The training stage in car plate identification for summarization is the same as the training stage of video summarization. The testing stage in car plate identification comprises test video preprocessing, detecting test car plate, SVM,

and LBP for optimization. Feature extraction using HOG feature, classification using Probabilistic Neural Network (PNN), to view the summary for a specific car.

The training process was supervised and the summarization type was dynamic because it's the suitable technique for surveillance video. The dataset that used in this thesis was the proposed dataset. The total time of local recorded videos is (19.5 minutes), (15.5 minutes) for training, and (4 minutes) for testing. The training samples were divided into (79.5%) for training and (20.5%) for testing. The proposed video summarization has got maximum accuracy of (83%) by using Viola-Jones and SVM with LBP. The informative frames retrieved from the original video were 17%. While video summary based on car plate identification achieves accuracy with (95%). The accuracy of the Viola-Jones object detection process for training 700 images is (97%). The accuracy of the SVM classifier is (99.6%).

CONTENTS

CHAPTER THREE

THE PROPOSED SYSTEM

CHAPTER FOUR

RESULTS AND TESTS

CHAPTER FIVE

CHALLENGES, CONCLUSIONS, AND SUGGESTIONS FOR FUTURE WORKS

LIST OF ABBREVIATIONS

Abbreviation	Meaning
AdaBoost	Adaptive Boosting
AI	Artificial Intelligence
ADL	Activity of Daily Living Dataset
BOW	Bag-Of-Words
CNN	Convolution Neural Networks
DBI	Davies-Bouldin Index
GMMs	Gaussian Mixture Model
HOG	Histogram of Oriented Gradients
ITS	Intelligent Transport System
IR	Information Rate
KL	Kullback-Leibler
LP	License Plate
LPR	License Plate Recognition
LSTMs	Long Short-Term Memory
LBP	Local Binary Pattern
ML	Machine Learning
NN	Neural Network
PNN	Probabilistic Neural Network
ROIs	Regions Of Interest
RR	Reduction Ratio
SCV	Sum Conditional Variance
SURF	Speeded Up Robust Features
SumMe	Summaries From User Video
SIFT	Scale-Invariant Feature Transform
SRD	Shot Reconstruction Degree
SVM	Support Vector Machine
TVSum	Title-based Video Summarization
VLPR	Vehicle License Plate Recognition

LIST OF FIGURES

LIST OF TABLES

LIST OF ALGORITHMS

Chapter One
General Introduction

Chapter One

General Introduction

1.1 Introduction

There is a huge amount of knowledge on the web where there's time-consuming when searching and browsing among extensive videos, therefore it's difficult to quickly get the specified event. The video summarization technique provides brief information about the whole video, briefly time, and makes browsing for large video faster. This makes video summarization more required and needed [1].

Summarization has been proposed initially for text data. The document summarization goal is creating an automatic summary for text almost like humans doing. The most ideal of a text should be identified and conveyed by the summary, and therefore the summary should be also precise and proper grammatically. Therefore, the non-important content and repetition must be avoided within the summary. Video and text summarization share many similarities and aim for similar goals [2].

A video summary is defined as a stream of still or moving pictures presenting the content of a video in such how that the relevant target is given brief knowledge while the fundamental message of the first video is preserved. There are two fundamental sorts of video abstraction techniques: The first is static video summarization which is additionally called representative frames, still-image abstracts, or static storyboards that summarizing the first video with a lot of data to a little number of frames without losing the rich information. While the second is dynamic video summarization also called dynamic video skimming, video skim, moving image abstract, or moving storyboard that summarizing the first video to video as short as possible that provides a global picture of the video. Most existing video summarization techniques are keyframe-based, i.e., several frames from the first videos are extracted to represent the entire video [3][4][5].

To summarize a video, most of the methods contain computing visual features from video frames, besides there are methods that consider the semantic meaning implied on videos to supply a more informative summary [6].

For years, vehicle number (license) plate recognition (VLPR) has been a subject of concern for several specialists, including those employed in the image processing field. Because of the growing number of cars, there is a need for an advanced traffic control device capable of recognizing, monitoring, and distinguishing a car that contravenes the law [7].

Such control includes License Plate (LP), area identification, character segmentation, and classification. There is no doubt that License Plate Recognition (LPR) systems need to react quickly enough to fulfill the requirements of Intelligent Transport Systems (ITS).

LPR systems would work so rapidly that no moving vehicle is missing [8]. One example of ITS is LPR, which can identify and differentiate vehicles, making it a very critical component of traffic systems [9]. Applications for the LPR program are traffic control, parking, and security. Advantages involve the availability of traffic jam information, and the speed of traffic and criminal activity are monitored [10].

Several novel features are extracted to characterize video boundaries, including cut, fade-in, dissolve, and dissolve for facilitating the understanding of content structure and domain rules of a video [11]. A video summary is either a static summary or a dynamic summary.

Machine learning and techniques are proved to achieve success for various image (video frame) analysis processes and object tracking [12]. A dynamic summary is a set of short video clips, joined in a sequence and played as a video. Therefore, this study uses machine learning (ML) for training and detecting car plates image to implement dynamic video summarization.

1.2 Related Work

Many types of research in the field of video summarization are developed. The present survey includes previous work related to this thesis:

- Nada Jasim Habeeb, et al., in 2016 [13], showed a surveillance video summarization method. This method assumed temporal differencing to obtain meaningful data from a large video stream. This technique used both histograms differentiate and Sum Conditional Variance (SCV) which were powerful against illumination alterations to obtain motion objects. The results showed that the presented technique was given better output in comparison with temporal differencing-based summarization methods with a compression ratio of 90%.
- Dipti Jadhav and Udhav Bhosle, 2017 [14], this paper suggest a methodology for video description based on the Speeded Up Robust Features (SURF). The authors also recommend an

approach based on graph theory to maximize the number of keyframes based on the objective function that the graph created by the optimized video description is a simple graph with a simple walk. The suggested algorithm is checked from the Open Video dataset on two separate videos, performance analysis, and subjective evaluation result 85%.

- Dong-Ju Jeong et al., in 2017 [15], proposed a two-step approach where the primary step skims a video. Therefore, the second step performs content-aware clustering with keyframe selection. The 1st step applying the spectral clustering technique with color histogram features. In the 2nd step, perform coarse temporal segmentation then apply refined clustering for each of the temporal segments, where each frame is represented by the sparse coding of Scale-Invariant Feature Transform (SIFT) features. Experiments result on videos with different lengths show that the resulting summaries closely follow the important contents of videos. UTE dataset results average F-frame measure 76.3%, ADL dataset results averaged F-frame measure 79.3%, and average precision 76.6%.
- Sinn Susan Thomas et al. in 2017 [16] explained how to utilize the best security camera description system. Besides that, the search time and proposing to turn content-based video retrieval issues into a content-based image retrieval concern. The query and the database matching using NN-classifier. The video was retrieved based on features such as Graph-Based Visual Saliency. This approach used Greedy Search Algorithm. This approach used two parameters to measure the performance of this system: The information rate

IR reflects the volume of information in the description assessing the efficiency of the condensed process. The reduction ratio RR is called the frame ratio summarizing the total frames in the recording, the average experimental result of this approach was 71% with IR=32% and RR=24%.

- Antti E. Ainasoja et al. in 2018 [17], This work proposes a simple but efficient dynamic extension of a video Bag-of-Words (BOW) system that provides over segmentation for keyframe pick at the same time as this technique, keyframes are selected from scenes that represent identically related material for scene detection. This research yielded a number of intriguing results. First, while area descriptors are mostly good at detecting scenes (visually identical content), optical flow (motion changes) offers stronger keyframes. Second, however, the appropriate criteria for motion descriptor-based keyframe selection vary from video to video, and the average output remains poor. To prevent more complicated computation, this paper proposes a human-in-the-loop phase in which the three best approaches yield user-privileged keyframes. Third, the human assistance and learning-free approach outperform learning-based approaches in terms of precision, and for some videos, it matches average human accuracy. The average result for different videos was egocentric videos 66%, moving videos 64%, static videos 59%.
- Madhav Datt and Jayanta Mukhopadhyay, in 2018 [18], presented a video summarization, by using convolutional neural networks (CNN) and bidirectional long short-term memory (LSTMs) to get deep features for frame

representation and to model variable-range temporal sequences. Further, they introduced a parameterized loss function minimizing (Kullback-Leibler divergence) KL-divergence between the Gaussian Mixture Model (GMMs) to find out relative orders of frame importance. This work expanded extensive evaluation on a lot of benchmarks (TvSum, SumMe, and YouTube) to determine the effectiveness of this model, Performance (F-score) of video summarization on the transfer supervised learning settings: SumMe 43.3%, TvSum 60.1%, YouTube 60.6%.

- Xin Ai et al. in 2018 [19], proposed an unsupervised video summarization method, which selects a group of highlight clips with self-consistency. Specifically, they proposed a consistent clip generation method, i.e. the cutting-merging adjusting scheme, by exploring the clip similarity and the local similarity. The consistent clips are obtained by merging similar clips iteratively and adjusting the boundaries of each consistent clip to remove the inconsistency of the boundaries between clips and logical events. Then, estimate the interest score of each consistent clip by computing the interestingness score of its frames, based on selecting the top important clips to generate a video summary. Experimental results presented using the SumMe dataset the relative was 76%.
- Muhammad Asim, Noor Almaadeed, et al, in 2018 [20], this paper presents a video description method for detecting shot boundaries based on the integration of color features derived video frame patches rather than a whole frame Per video shot is further broken down into sub-shots by measuring the structural similarity between frames to obtain a keyframe

from the most representative video shot sub-shots. Finally, the keyframes derived from and video shot's sub-shot are independently measured to eliminate redundant frames. The average experimental result extracted from the OpenVideo dataset was 67%.

- Muhammad Zeeshan Khan, Saira Jabeen et al. in 2019 [21], this paper presented a method in which first, determine the limits of the scene using movable characteristics. Subsequently, the data was passed to the proposed CNN architecture, which provides the frame-level value to each frame present in a specific scene. Experiments were carried out using the publicly available TVSUM50 dataset, the result was proposed (CNN+LSTM) 84 %.
- Seema Rani, Mukesh Kumar, in 2019 [22], a keyframe extraction method based on fusion from the visual characteristics is proposed in this research, which includes: correlation of RGB color channels, color histogram, mutual information, and inertia moments. As a clustering method, the Kohonen Self Organizing map is used to identify the most appropriate frames from the set of frames that come after fusion. Frames that are worthless are discarded and frames that have optimum Euclidean distance, with reselected as final keyframes in a cluster. The proposed technique is evaluated using degrees of fidelity and Shot Reconstruction Degree (SRD), with a YouTube video dataset. The average score for fidelity obtained using the proposed system was 64%.
- Debkumar Chowdhury, Souraneel Mandal et al. in 2019[23] proposed a method for license plate detection in three steps, proposed method mainly has three modules: 1) Detection of

license plate 2) Segmentation of Characters 3) Text Box Generation. The efficiency of this proposed system was 78.2%.

- Haibo Lin, Jianli Zhao et al in 2020[24] this paper proposed a method for license plate detection by, Firstly, the image preprocessing of the license plate includes graying and binarization. Then, the Sobel operator edge detection is performed according to the binarized license plate image. The Sobel operator has moderate sensitivity to the edge and is suitable for the extraction of the license plate edge. The experimental result was 90%.

1.3 Problem Statement

Nowadays, video represents one of the foremost objects utilized in social media, surveillance video, personal video… etc. Most of these videos might not have important information or might be repeated. This is going to add additional costs to the user because the video needs an outsized bandwidth to download or view it, in addition to an outsized space to store it. Solving the above problem is the main problem of this thesis.

1.4 Aim of Thesis

The objective of this thesis is to design and implement a package that can abstract a long surveillance video. The abstracted video also helps in the security aspect by detecting and identifying vehicle plate numbers. This work aims to build a model capable of training and detecting the passage of vehicles in long-sized videos, summarizing only specific areas of importance, and placing them in a brief video by applying a set of artificial intelligence. These techniques use the Viola-Jones algorithm for car plate detection by building models that depend on positive and negative samples. A set of different training models is applied and using the Support Vector Machine algorithm to optimize the car plate for the best result. A Probabilistic Neural Network (PNN) is used to test the car plate number.

1.5 Contribution

The contribution of this thesis is building a package for abstracting videos that can be used security manner. Also, the contribution of this work is represented by collecting a local dataset for training this system also, using the Viola-Jones algorithm for car plate detection, SVM for optimization manner, and using Probabilistic Neural Network (PNN) which is used to test the car plate number.

1.6 Layout of Thesis

The other chapters in this thesis are as follows:

- Chapter Tow “***Theoretical Background***”, presents a general overview of the methods used in this dissertation.
- Chapter three “***The Proposed System***”, presents in detail the proposed algorithms used to provide a video summary based on the features extracted and saved by machine learning techniques, the features obtained from the video itself.
- Chapter four “***Results and Tests***”, presents the outcome of subjective and objective measures of the proposed algorithms and therefore the time consuming for every processing step.
- Chapter five “***Challenges, Conclusions and suggestion for Future Works***”, present the conclusions drawn from this dissertation and provides suggestions for expansion this adds the future.