IJCSE International Journal of Computer Sciences and Engineering Open Access
Survey Paper Vol.-7, Special Issue, 5, March 2019 E-ISSN: 2347-2693

A Preliminary Investigation on a Novel Approach for Efficient and Effective Video Classification Model

M. Ramesh[1*], K. Mahesh[2]

[1,2] Department of Computer Applications, Alagappa University, Tamil Nadu, India

[*]*Corresponding Author: ramcsit@gmail.com, Mob.: +91 9884533306*

DOI: https://doi.org/10.26438/ijcse/v7si5.266269 | Available online at: www.ijcseonline.org

Abstract - In the recent ten to fifteen years web developers and web users expend more amount of time on images and videos. Since video is an admirable tool for delivering content, it has one of the major roles in human daily life. There are many kinds of videos available in real life and therefore we need an important tool to perform classification on video-based applications. Video classification and video content analysis is one of the ongoing research areas in the field of computer vision. The main goal of video classification is to help the viewers to find video of their own interest. We need a tool to classify the video with sky scramble accuracy. Therefore, we propose a model for video classification with several medium layers. This model takes video as an input passed through various layers and produce the video class label. The class label may be sports, movies, advertisement, cartoon, news etc.

Keywords - Video Classification, Keyframe, Video Frame, Background Subtraction.

I. INTRODUCTION

An image is a picture that may be created or copied and stored in an electronic device with electronic form. Image may have different components. The components are real life objects, human being, vehicle, plant and so on. Video is an electronic medium for recording, copying and telecasting. Video may have huge set of properties which may also be useful for managing video and its elements. It may be represented in the form of sequence of images or pictures. The images or pictures are called as frames. Frame is one of the still images in a video. The sequence of related images displaying may produce a video. In general video has 24 Frames Per Second (FPS) [1]. The quality of a video depends on the number of frames per seconds.

Video Classification:

Video classification can be simply treated as image classification. Unlike image classification, it has sequential frames input. Every frame per video can be labeled as the same video category [2]. There are numerous video classification methods and techniques. Few major techniques are, text-based approaches, audio-based approaches, video-based approaches and combination of the above. The essential term for video classification problems is to measure the accuracy and this may be calculated based on feature extraction and prediction.

II. RELATED WORK

One of the least common approaches in video classification is text-based approaches. F. Sebastiani [3] presented Text produced from a video falls into two categories: Viewable text placed on object in the video and text placed on screen. Text can be extracted from the object which is available in the video frame. The object may be name plate of house, shop name, sports person name displayed on their clothes and so on. Text may also be extracted from the screen, it may be score board displayed in bottom of the screen, and dialogue script displayed in bottom of the screen and so on. M.Ramesh, and K.Mahesh [2] presented if the text is not readable then the text is converted into readable text using OCR software. Features are extracted from the text and then given as input for any classification algorithms like SVM.

Audio based approach is one of the classification techniques. This is one of the best approaches to classify a video because it requires less computational resources compared with visual based approaches. Audio features require less space in case of storing feature in a database [4]. Features can be derived from either the time domain or frequency domain. To produce features from an audio signal, the signal is sampled at a certain rate (e.g., 22 050 Hz) [4].

Video is an electronic medium for recording, copying and telecasting. Video may have huge set of properties and this may also be useful for managing video and its elements.

Usually we call a video as a collection of images, which are known as frames. Number of frames within a single camera action is called a shot. A scene is one shot or number of shots [3]. Color based features and shot based features are used to classify the video. Many color spaces exist for representing the colors in a frame. Two of the most popular are the Red-Green-Blue (RGB) and Hue-Saturation-Value (HSV) color spaces. In the RGB color space, the color of each pixel is represented by some combination of the individual colors red, green and blue. In the HSV color space, colors are represented by hue (i.e., the wavelength of the color percept), saturation (i.e., the amount of white light present in the color), and value (also known as the brightness, value is the intensity of the color) [5].

Frame majority vote is also one of the video classification techniques. Video is predicted as a video category so that the video type can be obtained from the majority vote of frame predictions [1]. For example, we have 20 frames for a specific video, among these frames 12 frames are classified as news and 5 frames are classified as advertisement and 3 frames are classified as movie. Now using this method, the video has been labeled as news.

Object based classification is one of the complicated methods to classify the video. Tracking and detecting the object is more crucial work. Object tracking is one of the complicated research areas to track the object from the given video [1].Tracking a car from the car race sports video is an example for object based tracking. Once the object has been tracked the features are extracted and given for trained CNN.

III. PROPOSED MODEL

We have proposed a model for video classification with several medium layers. This model takes video as an input passed through various layers and produce the video class label. The class label may be sports, movies, advertisement, cartoon, news, movie teasers, cartoons, animated movies, short films and so on. The proposed model takes the following steps:

- Reads video as an input
- The input video is split into frames
- Identify frames and key frames
- Then it is pre-processed
- Background Subtraction
- Morphological processing
- Feature Extraction
- Produce class labels

The following diagram explores the working model of video classification.

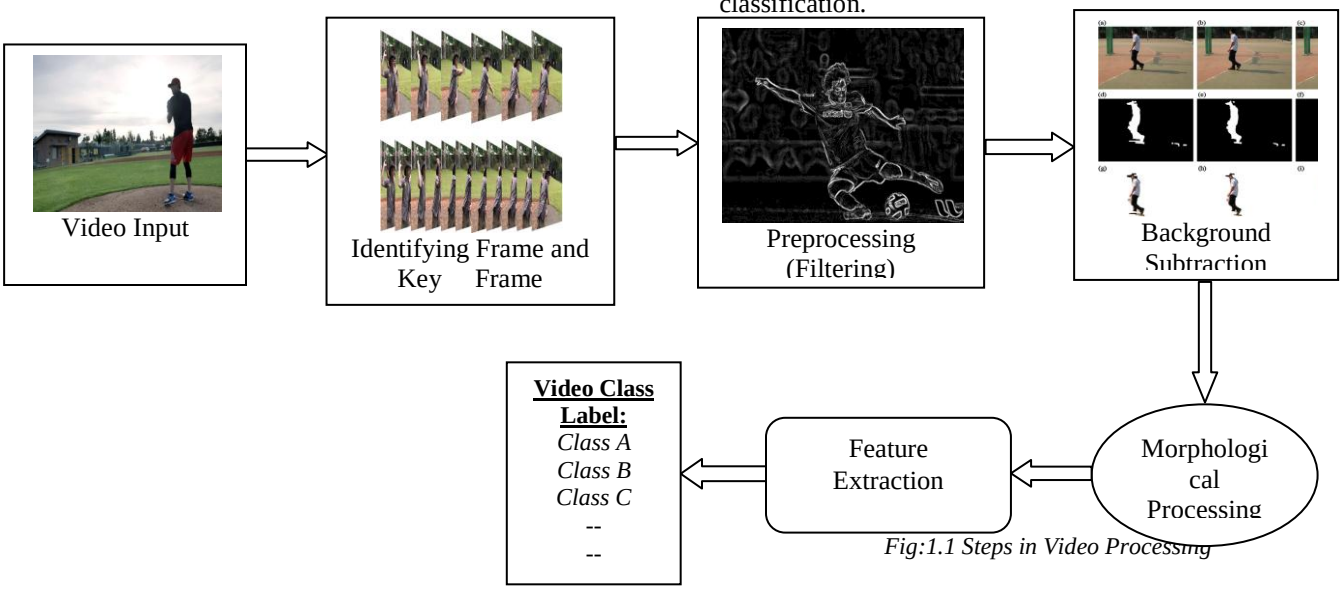

Fig:1.1 Steps in Video Processing

i. Video Input:

The proposed model takes video as an input and pass it to the next layer. The video may be in any of the popular file formats like avi(Audio Video Interleaved) , flv (Flash Video Format), wmv (Windows Media Video) , mov (Apple QuickTime Movie) , mp4 (Moving Pictures Expert Group 4). The input videos include walk, run, jump, jacking and car race action [6]. There are several properties for a video. The most common properties that we should know before start use the model are frame rate, resolution, width, height, interlace and metadata.

ii. Identifying frame and Keyframe:

Video is an electronic medium for recording, copying and telecasting and this may be represented in the form of sequence of images or pictures. The images or pictures are called as frames. Frame is one of the still images in a video. The sequence of related images displaying may produce a video. This state may include the following two sub states.

- Frame Selection
- Frame Elimination

As we discussed video is composed of sequence of frames. So single video may have greater number of frames and it depends on the length of the video. From these frames we need to select a set of frames and eliminate few of the frames based on the application where we need to implement. This process is known as frame selection and frame elimination respectively.

iii. Pre-processing:

The main purpose of pre-processing is an improvement of the frame data that suppresses unwanted distortions or enhances some frame features important for further processing. The common steps for pre-processing are

- Input frame
- Gray Conversion
- Filtering
- Binarization

Noise elimination, edge detection and shape reinforcement are the few of the examples for pre-processing. Gaussian filter is one of the most common techniques to remove the noise in the input frame.

iv. Background Subtraction:

Generally, frame's (images) regions of interest (RoI) are objects (cars, text, ball, humans etc.) in its foreground. Background Subtraction is a widely used approach for detecting moving objects in frame sequence i.e videos. This is a very most trendy method of movement segmentation. In this state images are called as frames. Commonly the first frame is taken as a reference and each remaining frames of the sequence is compared to this first frame [2]. Differences of the frames are deeming as moving objects. Now moving objects are deem as foreground objects.

The current frame and the background image are differenced to create the binary background object (may be ball, car, human, apple etc…).

$$o(u,v) = i(x,y) - b(x,y) \quad (1)$$

where i(x,y) is the input frame pixel, b(x,y) is background frame pixel and o(u,v) is the differenced object pixel.

v. Morphological Processing:

Morphological processing is a collection of non-linear operations related to the shape or morphology of features in an image. The main goal of applying morphological processing is to remove noise from an image. There are three major operations in morphological processing.

- Erosion
- Dilation
- Closing

Erosion is used for removing the unwanted portions other than ROI (Regions of Interest), Erosion of a binary image f by a structuring elements (denoted $f \ominus s$) produces a new binary image $g = f \ominus s$. The dilation of an image f by a structuring element s (denoted $f \oplus s$) produces a new binary image $g = f \oplus s$ with ones in all locations (x,y). Finally morphological closing is used for filling the holes in the ROI [6].

vi. Feature Extraction:

Generally, Features are represented in a form of numerical information that is difficult to understand by human. Feature detection, extraction, and elimination are often combined to solve some common computer vision problems such as object detection and recognition, content-based image retrieval, face detection and recognition, image classification and so on. There are two types of features extracted from a frame (images). They are local and global. The following properties are considered before classifying the frames. Class informative: all objects of the same class must be represented by the same set of descriptors. Discriminative: Each class has its proper set of descriptors. Minimal: should not contain redundant descriptors [6].

vii. Classification:

This layer takes input as feature set and produce output as class label for the given video. Classification is the most crucial job in this model. Selection of classification algorithms is one of the most important factors for video classification. Classification problems are usually posed as a supervised learning problem, where a set of examples also called as training set [5]. Moving objects and its feature are extracted and then these features are compared to the set of labeled examples provided in the training set. The followings are the common classification algorithms for the given data set.

Linear Classifiers: Logistic Regression, Naive Bayes Classifier.

- Support Vector Machines.
- Decision Trees.
- Boosted Trees.
- Random Forest.
- Neural Networks.
- K - Nearest Neighbor.

IV. CONCLUSION AND FUTURE DIRECTION

In this paper we have given a theoretical foundation of a model for video classification with several medium layers.

This model takes video as an input passed through various layers and produce the video class label. The class label may be sports, movies, advertisement, cartoon, news etc. The main goal of video classification is to help the viewers to find video of their own interest. The focus of this paper is to propose a model to classify the video with sky scramble accuracy. This proposed model is yet to be implemented. Video classification and video content analysis is one of the ongoing research areas in the field of computer vision. As an extended work, this proposed model can be implemented using any tool like matlab, R, and python. The goal is to provide a model with more accuracy with high level efficiency for Video Classification.

REFERENCES

[1] M. Ramesh1*, K. Mahesh2, "*Multidimentional View of Automatic Video Classification : An Elucidation*", International Journal of Computer Sciences and Engineering, Vol-6, Special Issue-4, May 2018.

[2] Jiajun Sun, Jing Wang, Ting-chun Yeh, "*Video understanding: From video classification to captioning*".

[3] Nirav Bhatt, - "*A survey on video classification techniques‖, International Journal of Emerging Technologies and Innovative Research*", March 2015, Volume 2, Issue 3.

[4] Darin Brezeale and Diane J. Cook, - "*Automatic Video Classification: A Survey of Literature*", IEEE Transactions on systems, man, and cybernetics—part c: applications and reviews, vol. 38, no. 3, may 2008.

[5] A.D.Bimbo, "*Visual information retrieval, San Francisco*", CA: Morgan kaufman, 1999.

[6] Insaf Setitra, "*Object Classification in Videos—An Overview*", Journal of Automation and Control Engineering, Vol. 1, No. 2, June 2013

[7] M. Ramesh1*, K. Mahesh2, ," *Significance of various Video Classification Techniques and Methods: A Retrospective*",International Journal of Pure and Applied Mathematics, Volume 118 No. 8 2018, 523-526.

[8] S.Maheshwari, P. Arockia Jansi Rani, "*Human Action Recognition System Based on Silhouette*", International Journal of Computer and Information Engineering, VOl. 9, No:110, 2015.

[9] Darin Brezeale and Diane J. Cook, ―"*Automatic Video Classification: A Survey of Literature‖, IEEE Transactions on systems, man, and cybernetics—part c: applications and reviews*", vol. 38, no. 3, may 2008.

[10] Reza Fuad Rachmadi†∗ , Keiichi Uchimura†, and Gou Koutaki, - "*Video classification using compacted dataset based on selected keyframe"*, Proceedings of the International Conference,2016 IEEE Region.

[11] Shankar, K., K. Mahesh, and K. Kuppusamy. – "*Analyzing Image Quality via Color Spaces"* image. In 1.1 (2014).

[12] Mittal C. Darji1, Dipti Mathpal2, - " *A Review of Video Classification Techniques",* International Research Journal of Engineering and Technology (IRJET), Volume: 04 Issue: 06 | June -2017.

Authors Profile

Mr. M.Ramesh pursed Bachelor of Science in Computer Science, Master of Science in Computer Science and Information Technology and M.Phil in Computer Science from Madurai Kamaraj University, Tamil Nadu, India. He is currently pursuing Ph.D (part time) in Algappa University and currently working as Assistant Professor in Depatment of Computer Science, Faculty of Science and Humanities, SRM IST, Chennai since 2011. He has more than 11 years of teaching experience.

Dr. K Mahesh pursed Master of Computer Applications, M.Phil in Computer Science and Ph.D in Computer Science. He is currently working as Professor in Department of Computer Applications in Alagappa University, Tamil Nadu, India. He has published 45 International journals, 9 International Conference papers, 3 National Journals and 23 National Conferences. He has completed funded research project titled "Collaborative Directed Basic Research in Smart and Secure Environment" from july 2007 to august 2012. He is a member of International Association of Engineers (IAENG). He is Reviewer, ICTACT Journal on Image and Video Processing (IJIVP) Publisher: ICT Academy of Tamil Nadu and also Reviewer, International Journal of Advanced Research Trends in Engineering and Technology (IJARTET) Publisher: ACE Publishers. He has Presented Key Note Address in the National Seminar on Current Trends in Computing Technologies organized by the PG Department of Computer Science, Government Arts College for women, Ramanathapuram, Feb 28, 2015 and he aslo presented Presented Key Note Address in the Intercollegiate Meet (TECHNO'15) organized by the PG Department of Computer Science, Idhaya College For Women, Sarugani, Sep 9th ,2015. He has 26 years of teaching experience and 9 years of Research Experience.

