IJCSE International Journal of Computer Sciences and Engineering Open Access

Research Paper Vol-6, Special Issue-4, May 2018 E-ISSN: 2347-2693

Multidimentional View of Automatic Video Classification : An Elucidation

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

[1*] Department of Computer Applications, Alagappa University, Karaikudi, India
[2.] Department of Computer Applications, Alagappa University, Karaikudi, India

**Corresponding Author: ramcsit@gmail.com, Mob.: +91-98845 33306*

Available online at: www.ijcseonline.org

Abstract — Media is one of the foremost roles in human daily life activity. Multimedia is the integration of multiple forms of media, which includes text, image, audio, and video. Most of the people are always working with their Personal Digital Assistant (PDA) that provides computing, information storage and retrieval capabilities for personal or business use. Images and videos engage more space than other kinds of data on their PDA or electronic device. There are many kinds of videos available in day to day life, so we need an efficient tool to classify the videos with sky-scraping accuracy. The main goal of video classification is to help the people to find video of their interest. In this paper we study multi dimensional view of video classification methods and techniques, compare them and also conclude with opinion for further research.

Keywords—

I. INTRODUCTION

Images and videos are shared all the way through social Medias such as facebook, youtube, linkedIn etc in day to day life. As per the survey there are 2 billion monthly active users (June 2017). Especially video has one of the major part in humans routine life.

A. *Video:*

Video is an electronic medium for the recording, copying, playback, broadcasting, and display of moving visual media. In general video is nothing but sequence of images, which may be sometimes call it as frames. Video has different set of properties such as size, format, bit rate, frame rate etc…

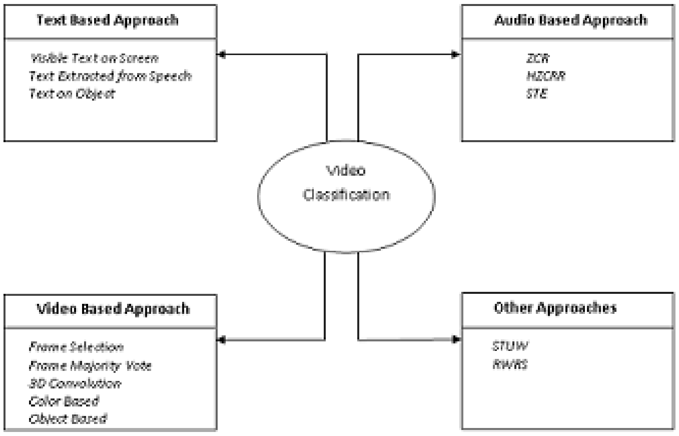
B. *Frames:*

A frame is one of the many still images which may produce the complete moving picture. In general video has 24 frames per second (fps). Video size can be calculate based on the following properties such as frame, image size, color and so on. The quality of video is based on the number or frames per second used in a video.

C. *Video Classification:*

There are huge volumes of video data available in day to day life for personal and business use. So we need an automatic video classification tool to categorize the video for processing it. The main intention of video classification is to categorize the video whether is coming under which variety. There are few major groups are there such as movies, short films, sports, cartoons and animation movies, news, weather report etc… Video classifications can be done by the following techniques [1][2].

1. Text based approaches
    a. Visible text on screen
    b. Text Extracted from speech
    c. Text on object
2. Audio based approaches
3. Video based approaches
    a. Frame Selection
    b. Frame Majority vote
    c. Temporal Feature Pooling (TFP)
    d. 3D Convolution (C3D)
    e. Color based
    f. Object based
4. Other approaches



D. Applications of video classification:

Many videos are available in day to day life. So we need an efficient tool to search a specific video from the video database. The main goal of video classification is to help the people to find video of their interest from the internet. There are huge varieties of applications available. Here we have mentioned few examples. In social web site we are receiving unwanted video that we don't want at most. A high jump sport video consists of two different actions, running and high jump, which shared with other videos such as running or hurdling sport video.

II. EARLY TECHNIQUES AND METHODS

In this section, the author describes the previous research works in the form of title, problem statement, objectives, not repeat the information discussed in Introduction [2].

A. Text Based Approach:

One of the video classification approaches is text based approach, which is least common approach. There are four common methods for text based classification visible text on screen, text extracted from speech [3], text on object and closed caption.

a. Visible Text on screen:

This is one of the approaches to classify video on text based approach. Text displayed on the bottom of the screen relevant to conversations of the frames in a video. 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 tool like SVM. Finally machine will label the video as specific category.

b. Text Extracted from speech:

This is also one the approaches to classify video on text based approach. Sample audios are separated from the given video and it convert into text using speech recognizing software. Based on the text features video can be classified.

c. Text on Object:

One more method to classify the video based on text displayed on object. Texts are displayed on object are captured and converted into viewable text by using OCR (Optical Character Recognition) tool. Example for text on object is vehicle registration number plate, Building name, city name and so on.

B. Audio Based Approach:

Audio based approach needs little computation and processing resource compared with video based approach. This method is tricky to distinguish with multiple sound audio file. There is multiple level of audio based feature. Time and frequency domain features are the low level features [4]. Volume standard deviations and volume dynamic range can be used to detect whether the video has constant level of noise. Zero crossing rate (ZCR) is the number of signal amplitude sign changes in the current frame. If the loudness and ZCR are both below thresholds, then this frame may represent silence. [4] Bandwidth is a measure of the frequency range of a signal. Some types of sounds have more narrow frequency ranges than others. Speech typically has a lower bandwidth than music.

C. Frame Based Approaches:

A frame is one of the many still images which may produce the complete moving picture. A frame is sometimes called as image. Frame based video classification itself is one of the major research area. There are few methods to classify the video based on frame and frame based properties [2][5].

a. Frame selection:

Selecting the frame itself is one of the crucial works to be done before classification of given video. There are huge amount of frames for a single video. From these frames we have to select the frame by implementing some classifier algorithms. Depending upon the algorithm the classification accuracy may increase or decrease and time complexity may also increase and decrease. Once the frames are selected from the video these set of frames are called as keyframes. After creation of the keyframe, features are extracted from the keyframe and give as input for any classifier.

b. Frame Majority Vote:

Video classification is similar to image classification. Every frame in a single video can be labeled as same category and features are extracted. Then train any ANN or CNN classifier using the dataset of the frames. In test mode, every frame per video is predicted as a video category so that the video type can be obtained from the majority vote of frame predictions. For example there are 30 frames in a video among these 25 frames are labeled as spots and the remaining 5 frames are labeled as advertisement. Now this video is classified as sports category since it has 25 majority frames labeled as sports.

c. 3D Convolution Approach:

C3D is one of the simplest and effective approaches, used to deep 3-dimensional convolution neural networks trained on a large scale video dataset. C3D operates on stacked video frames and extends the original 2D convolution kernel and 2D pooling kernel into 3D kernel to capture both the spatial and temporal information. [Class-caption] However, training a 3D CNN is very time consuming and the spatial temporal structure in videos may be too complex to capture.

d. Color Based:

Video is nothing but sequence of related images sometimes called as frames. Frames are composed of lines. Each line is sampled to create number of pixels per line. Resolution of the frame is dependent on number of lines per frames and number of pixels per line. RGB, CMYK are two common color spaces . The color pixel is represented by combination of the individual colours red, green and blue in some amount. In the HSV colour space, colours are represented by hue (i.e., the wavelength of the colour percept), saturation (i.e., the amount of white light present in the colour), and value (also known as the brightness, value is the intensity of the colour) [5].

e. Object Based Approach:

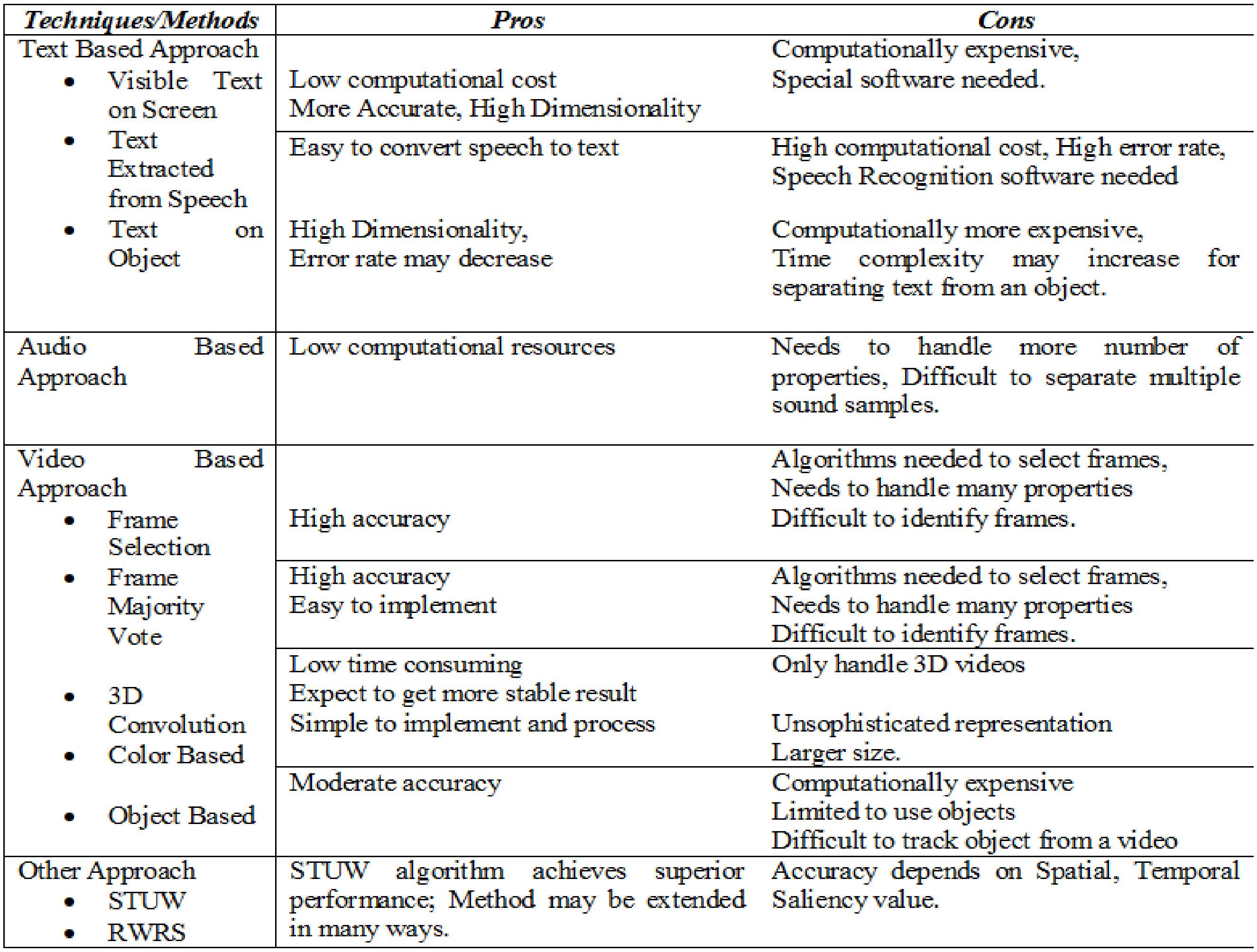
This is one of the complicated methods to classify the video. Tracking and detecting the object is more crucial work. Object tracking is one of the complicated research area to track the object from the given video. Tracking a ball from the foot ball sports video for further process is an example for object based. Once the object has been tracked the features are extracted and given to trained CNN.

D. Other Approach:

- STUW – Spatiotemporal Uncertainty Weighting [6]
- RWRS – Random Walk with Restart Saliency [6]

The above methods and approaches are used to classify the given video based on the category. Based on the approaches and methods features are extracted from the given video and then classified into specific genre. Following categories are the most frequently used genre.

i. Movie
ii. Movie Trailer
iii. News
iv. Sports
v. Advertisements
vi. Short Films
vii. Animated Movies
viii. Cartoons, etc…



Techniques/Methods	***Pros***	***Cons***
Text Based Approach • Visible Text on Screen	Low computational cost More Accurate, High Dimensionality	Computationally expensive, Special software needed.
• Text Extracted from Speech	Easy to convert speech to text	High computational cost, High error rate, Speech Recognition software needed
• Text on Object	High Dimensionality, Error rate may decrease	Computationally more expensive, Time complexity may increase for separating text from an object.
Audio Based Approach	Low computational resources	Needs to handle more number of properties, Difficult to separate multiple sound samples.
Video Based Approach • Frame Selection	High accuracy	Algorithms needed to select frames, Needs to handle many properties Difficult to identify frames.
• Frame Majority Vote	High accuracy Easy to implement	Algorithms needed to select frames, Needs to handle many properties Difficult to identify frames.
• 3D Convolution • Color Based	Low time consuming Expect to get more stable result Simple to implement and process	Only handle 3D videos Unsophisticated representation Larger size.
• Object Based	Moderate accuracy	Computationally expensive Limited to use objects Difficult to track object from a video
Other Approach • STUW • RWRS	STUW algorithm achieves superior performance; Method may be extended in many ways.	Accuracy depends on Spatial, Temporal Saliency value.

III. CONCLUSION AND FUTURE DIRECTION

There are huge volumes of video data available in day to day life for personal and business use. So we need an automatic video classification tool to categorize the video for further processing. We have reviewed various video classification methods and techniques and explored more ideas on it. Most of the methods and techniques are compared along with features and their pros and cons are tabulated. Each and every technique use to handle different set of features to classify the given video in a pre defined specific genre. All the features are drawn from three major modalities; they are text, audio and video. The main goal of video classification is to classify the data set on a specific category with high accuracy in order to help the people to find video of their interest. With this details we conclude frame selection and feature extractions is one of the major role to classify the video with sky-scraping accuracy.

To improve the accuracy of video classification we need to draw on some better method to select the frames and its features. Increasing accuracy and efficiency of video classification is our extended work along with existing methods and techniques like keyframe extraction and feature selection. To have better performance, extracting more frames with some diverse methods and techniques from video are likely to be helpful.

REFERENCES

[1] 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.

[2] Jiajun Sun, Jing Wang, Ting-chun Yeh, - "*Video Understanding: From Video Classification to Captioning*", 2017.

[3] 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.

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

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

[6] Hansang Kim, Youngbae Kim, Jae-Young Sim, and Chang-Su Kim, ―"*Spatiotemporal Saliency Detection for Video Sequences Based on Random Walk With Restart"*, IEEE transactions on image processing, vol. 24, no. 8, august 2015.

[7] 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.

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

[9] Shankar, K., K. Mahesh, and K. Kuppusamy. "*Video Segmentation on 2D Images with 3D Effect.*" International Journal of Computer Applications 43.8 (2012): 1-4.

[10] 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

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 10 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 25 years of teaching experience and 9 years of Research Experience.