Proceedings of the 2nd International Conference on Software Engineering and Machine Learning
DOI: 10.54254/2755-2721/73/20240378

Content-based video adaptation algorithm

Shunzhe Yang

University of California, San Diego

shy033@ucsd.edu

Abstract. Video adaptation is crucial in today's digital landscape, as largely quantity of digital contents becomes valuable source. In this paper, a novel content-based video adaptation algorithm is proposed for video customizing process. This algorithm contains two major parts, content detection and framerate adaption. Content detection utilizes face recognition, object detection and static detection to understand the content info of the processed video. Using that info, framerate adaptation is utilized to control the framerate, achieving the video customization and only compressed the uninterested and unimportant content. Analysis show that the method could be implemented in various fields such as sports and film industries where digital contents have important role in decision making.

Keywords: Video Adaptation, Video Analysis, face recognition, object detection.

1. Introduction

Videos have become one of the most prevalent forms of media in today's digital landscape, documenting countless events, experiences, and interactions. As the volume, quality, and length of content rise, finding relevant information with efficiency has become a crucial task for everyone. This challenge becomes even more pronounced when the objective is to recognize specific individuals in a video stream. As a result, new technique should be introduced to this landscape to improve the efficiency and variety of video analysis. A personalized video filtering interface best accomplishes this demand since it makes efficient retrieval of key contents.[1] Even though such technologies face challenges to accomplish, advancing video analysis technologies have immense significance in a wide range of applications, offering opportunities to unlock valuable insights and enhance various aspects of digital contents.

Video adaptation refers to the process of modifying video content to suit different conditions in order to provide the best viewing experience. It involves making adjustments to various aspects of the video, such as resolution, bitrate, format, and encoding.As devices used to play videos vary in size and pixels, it is important to adapt the difference for videos to be played as intended. Higher-resolution videos may be suitable for devices with large, high-definition screens, while lower-resolution versions are more appropriate for smaller screens to maintain clarity without overloading network bandwidth. [12] Yet finding the optimal resolution based on difference in devices and conditions could be challenging and researchers have shown interest in this field of study. One strategy regarding this issue is to create video quality prediction model based on frame differences. Based on space-time displaced frames difference, a model is constructed to predict how people’s actual perception of video quality is affected with different combination of displaced frames.[2] Another solution involving prediction with model is built on an adaptive super-resolution framework. The framework used a reinforcement learning model to

DOI: 10.54254/2755-2721/73/20240378

integrate both super-resolution technique(VSR) and video streaming strategy(video reconstruction). The framework considers "client-end computing capabilities" as one feature when training the model and combines that with objective perception of video quality to predict the best resolution for reconstruction.[3]

Another crucial aspects in video adaptation is adjusting the video's bitrate, which determines the amount of data transmitted per unit of time. During moments of limited network connectivity, lowering the bitrate can prevent buffering and interrupted playback. Alternatively, higher bitrates can be applied when network conditions are favorable, delivering more detailed experience. By carefully calibrating the bitrate, video content can be delivered efficiently while maintaining a balance between visual quality and bandwidth consumption.[4] Similar question and solution are raised as that of resolution: researchers wonder which bitrate and transmission would best present the video content to the audience and they do so by proposing various models. These models could be classified into categories based on UHD video signals: PSNR, SSIM, VMAF, and J.342. After plotting these data into regression, the researchers could determine the suitable bitrates and transmission for a given UHD signal.[5] Another solution to bitrate variation is proposed with a model of satisfied-user-ratio(SUR). This model utilizes VideoSet, a large-scale data set released in 2020, along with Gaussian Process Regression to make the prediction more practical. Just like other approaches, this method takes viewers' experience into account of features when training. Yet the prediction is made in an alternative way: a certain bitrate is suggested by the model given a user-desired SUR(expected quality of experience). This approach has the most practical use as it has very little restriction on operating video adaptation individually.[6]

Furthermore, video adaptation encompasses the transformation of the video's format and encoding to cater to requirements of different devices. This requires converting videos into widely supported formats, such as MP4 or WebM, to ensure broad compatibility. Additionally, encoding techniques, such as using efficient codecs like H.264 or H.265, can be employed to minimize file sizes without compromising visual quality, enhancing both streaming performance and storage efficiency. An additional improvement on these techniques is HEVC(High Efficiency Video Coding) which provides higher compression ratio while preserving most of video quality.[8] Yet it demands considerable computing capabilities of devices and an estimator-allocator framework is put in place to solve the issue. The framework has a resource estimator that determines the policies to maximize resource utilization and encoding time. The allocator executes according to the task scheduled by the estimator and it also has a monitor tracking down the states of progress such as hardware usage and internet transmission, making responsive changes in policy possible. This technique helps audience to take advantage of higher experience with video contents with limited hardware and internet conditions.[7]

Current video adaptation focus more on the hardware limitation to adapt its content. This paper focus on another import part from the perspective of human being. People shows different interest on different content of a same video. Playing the video at the same frame-rate will cause people feel boring in the period of content they are not interested. Hence, this paper proposed a new algorithm called as content-based video adaptation algorithm. The core idea is to compress the content that the users has no interests and only play the content the user focus more. To achieve this, this algorithm uses face recognition, object detection and static detection to learn the content info from the video. Then, this algorithm control the framerate based on the content info and user's interest. This algorithm shows potential in social media, entertainment, and TV shows.

2. Content-based Video adaptation algorithm

As shown in Figure 1, this paper contains the following parts: face recognition, object detection, static detection, static removal and object-based frame rate adaptation.

DOI: 10.54254/2755-2721/73/20240378

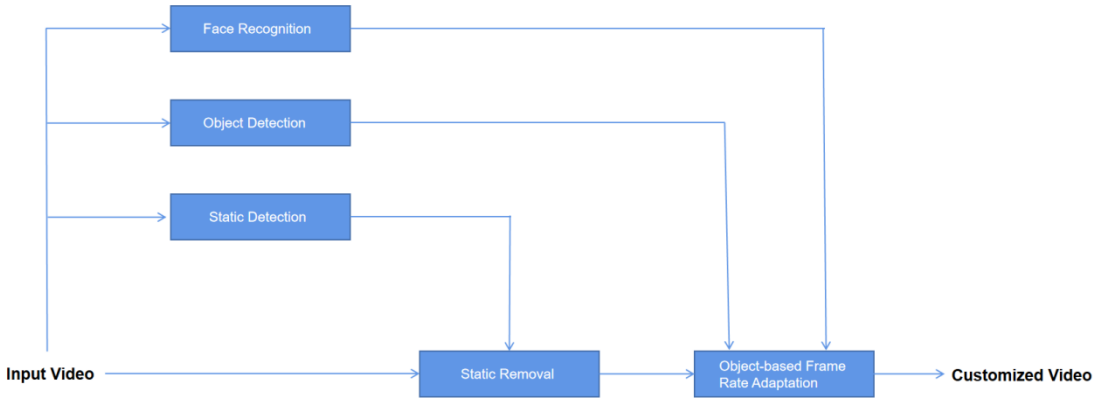

Figure 1. Content-based video adaptation algorithm

2.1. Static Detection

A proposed method is static segments detection and classification which is based on video temporal activity. A sequence of frames is considered "static" if their difference in consecutive temporal activity doesn't surpass a given threshold. The threshold value would be customized and it would not be zero due to the existence of noise. Nevertheless, there could be potential false detection because of marginal temporal activities such as small object with little movement. The proposed way to deal with this issue is to set another threshold that compares the difference between the first and the current frame of the sequence. By comparing frames with greater temporal distance, false detection regarding small motion could be reduced.[9] An additional feature to fortify the accuracy of detection is the Y Color Model which converts input videos from RGB to YUV, capturing significant color changes between frames. This technique works efficiently in videos with significant hue and lighting contrast.[10]

2.2. Static Removal

Using the proposed methods of static detection, frames that aren't changing are identified and set aside for removal. Conversely, frames displaying significant dissimilarity are flagged as dynamic and earmarked for preservation. Then, the selected dynamic frames are exported, creating a condensed video sequence that exclusively captures the moments of interest. By excluding irrelevant information, computational resources are conserved during real-time analysis, highlighting crucial temporal events. This approach is advantageous for applications like motion detection and activity recognition, as it ensures that the computational effort is concentrated on frames that contribute meaningfully to the video's narrative, thus optimizing the overall video analysis.

2.3. Object Detection

Object detection methods include image segmentation, motion segmentation, and change segmentation algorithms. Change segmentation-based algorithms excel in responding to upcoming frames and allowing larger nonrigid motion. One method is to process the video as each frame is filtered into: high frequency vertical information, high frequency horizontal information, high frequency diagonal information, and low frequency information. The difference between frames is calculated with a "wavelet domain mask" that compares it with a threshold value calculated by significance test. The domain masks will be used later to map out the extracted image. The Canny Operator is used for edge detection. Pixels in a frame is assigned a static index. If the pixel changes, its static index becomes zero; if a pixel has high static index, then it is likely a background pixel which could be used to determine the edges. The edge pixels are then processed for moving edge detection. They are mapped according to edge changes and then being compared with the "still" edges and finally the shape of edges is described with a set that contains pixels that are from both moving and still edges.[11]

DOI: 10.54254/2755-2721/73/20240378

2.4. Face Recognition

The proposed face recognition system utilizes three verification modules: face skin, face symmetry, and eye template, eliminating non-face objects and training a classifier using three individual Artificial Neural Networks.

For Face Skin Verification Module, regardless of ethnicity, human skin has distinct spectra in the color space. Using the k-nearest neighbor (k-nn) cluster algorithm, 1024 facial images are categorized into four groups of color space. For the purpose of face verification, two color spaces, (r, g, b) and (H, S, V), are selected due to their consistent performance and reduced vulnerability to factors like illumination, intensity variations, and partial light obstruction.

The Face Symmetric Verification Module uses a Y grayscale image to verify face symmetry, segmenting the face area using color and width information, dividing it vertically and evenly into left and right sub-blocks, calculating histograms, and calculating the Symmetric Similarity Coefficient.

After passing the previous two verification, the face area image is processed for the eye template verification module. A novel classifier combination scheme is proposed for video-based face recognition, aiming for low error rate and high recognition rate. The recognition criteria are set through genetic evolution, ensuring optimal weights. A recognition result is accepted if three ANN classifiers vote for the same person. Overall, this novel face detection system incorporates three hybrid feature sets and an ensemble classification scheme, enhancing overall recognition rate and reliability with adjustable reliability of frontal face detection by setting verification thresholds.[13]

2.5. Object-based Frame Rate Adaptation

One method of object-based frame rate adaptation involves utilizing real-time object tracking and prioritization algorithms. In this approach, the video playback system identifies and tracks individual objects or regions of interest within the video frame. By continuously monitoring the importance and complexity of each tracked object, the system dynamically adjusts the frame rate for these objects accordingly. Critical or visually significant elements are allocated a higher frame rate to ensure smoother and more detailed rendering, while less essential or static elements may receive a lower frame rate, conserving computational resources. Combining with the previous filters, the object detection and face recognition process would assign levels of significance to identified objects while the static detection and removal process cuts out “uninterested” frames. The lasting frames would be played at designated frame rate according to their level of importance.[14]

3. Analysis

There are many potential application scenaries of this algorithm, benefiting from its ability to efficiently filter through videos and extract segments featuring a particular person

In sports industry nowadays, scouts and managers have to examine players with more support from data than their guts. Yet sports that emphasize on collective success such as basketball and football have more complex standards than those that emphasize individual performance. Some players have magnificent impact when they have possessions while others contribute to different aspects of the game without being captured by the broadcast camera. A video filtering interface would come in handy for analysts to examine a player as a whole as it tracks the player’s performance for the entire game given sources of video from all cameras. Although it’s not hard to track a player for one game, professional club often needs to study lots of players for seasons before they make their decisions since players are expensive assets in the industry. With more data available, the teams would improve efficiency analyzing and make better decisions when purchasing their desirable players. The illusive figure is shown in Figure 2.

Another application of such interface would be in film industry. Similar to sports industry, when selecting actors, it would improve the director’s the decision if an actor’s overall ability is addressed rather than choosing with only a few highlight clips in hand. Yet an actor could participate in so many lengthy work and it is not possible for someone to watch them all to examine the actor’s performance. With the assistance of video filtering, the computer can pick out all the scenes with the actor involved

DOI: 10.54254/2755-2721/73/20240378

which would boost the efficiency of examination. Furthermore, a recommendation system could be established with the selected segments as the actors could be scored based on their performance. A customized filter interface also allows more detailed tracking. For instance, the scenes which the actor only shows up a little will have marginal effect on the evaluation so they could be skipped for efficiency. With the assistance of video filtering, the director will have more data that help to make the decision on which actor fills the role the best.

Personalized Video Filtering Interface helps to boost efficiency as its automation helps to analyze large quantity of videos effortlessly. Its best advantage is to make connection between videos with the data gathered in each segment. As the society's demand on quality data keeps increasing, such customized interface could be their new source of data producer. By utilizing the these interfaces, researchers, industry professionals, and content creators can make informed decisions and develop strategies for the needs and expectations of their work.

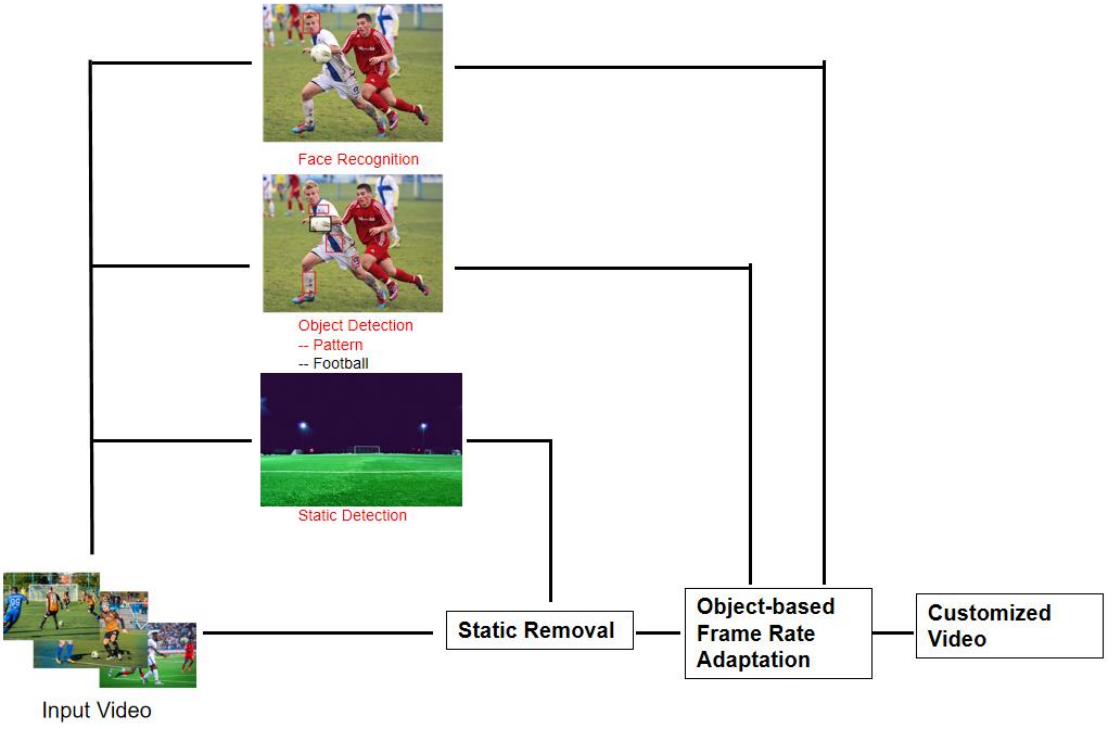

Figure 2. Sample Application of Detection-Adaptation Framework

4. Conclusion

In this paper, the content-based video adaptation algorithm is proposed to do the content analysis and frame rate adaptation. Using static detection, face recognition, and object detection, the algorithm is able to learn the existence of particular person, the interested object or useless static frame through the video frames. Using frame rate adaptation, this algorithm is able to compress different frames with different ratio. This will ensure the majority of video content will be in the user's interests. This has been analyzed to show big industrial values in sports and film.

References

[1] Zhang W, Wang Y, Chen H and Wei X 2019 China. An efficient personalized video recommendation algorithm based on mixed mode *IEEE IUCC* and *DSCI* and *SmartCNS* pp. 367-73

DOI: 10.54254/2755-2721/73/20240378

[2] Lee D, Ko H, Kim J and Bovik A 2020 Italy. Video quality model for space-time resolution adaptation *IEEE 4th International Conference on IPAS* pp. 34-9
[3] Zhang Y et al. 2020 Canada. Improving quality of experience by adaptive video streaming with super-resolution *IEEE Conf. on Comp. Commu.* pp. 1957-66
[4] Javadtalab A, Omidyeganeh M, Shirmohammadi S and Hosseini M 2017 Taiwan. A bitrate-conservative fast-adjusting rate controller for video *IEEE ISM* pp. 338-41
[5] Lee C, Woo S and Baek S 2019 Greece. Bitrate and transmission resolution determination based on perceptual video quality *10th Inter. Conf. on IISA* pp. 1-6
[6] Zhang X, Yang C, Wang H, Xu W and Kuo C 2020. Satisfied-user-ratio modeling for compressed video ieee transactions on image processing vol. 29 pp. 3777-89
[7] Lv H et al. 2014 Australia. A resolution-adaptive interpolation filter for video codec *IEEE ISCAS* Australia pp. 542-5
[8] Rasch J et al. 2018 Greece. A signal adaptive diffusion filter for video coding using directional total variation 25th *IEEE ICIP* pp. 2570-4
[9] Almeida R and Queluz M 2016 Morocco. Automatic detection & classification of static video segments *5th ICMCS* pp. 93-8
[10] Surit S and Chatwiriya W 2011 South Korea. Forest fire smoke detection in video based on digital image processing approach with static and dynamic characteristic analysis *First ACIS/JNU Inter. Conf. on Comp., Net., Sys. and Ind. Eng.* pp. 35-9
[11] Zhang X and Zhao R 2006 China. Automatic video object segmentation using wavelet transform and moving edge detection *Inter. Conf. on Mach. Lea. and Cyb.* pp. 3929-33
[12] Ren S, He K, Girshick R and Sun J 2017. Faster R-CNN: Towards real-time object detection with region proposal networks *IEEE Tran. on Pat. Ana. & Mach. Intel.* vol. 39 no. 06 pp. 1137-49
[13] Zhang P 2008 Washington DC. A video-based face detection and recognition system using cascade face verification modules *37th IEEE App. Ima. Pat. Rec. Work.* pp. 1-8
[14] Hou J, Wan S and Yang F 2010 Hong Kong, China. Frame rate adaptive rate model for video rate control *Inter. Conf. on Mul. Commu.* pp. 226-9