YouTube Video Categorization Using Moviebarcode

Recep Erol, Rick Rejeleene, Richard Young, Thomas Marcoux, Muhammad Nihal Hussain, and Nitin Agarwal
Collaboratorium for Social Media and Online Behavioral Studies (COSMOS),
University of Arkansas at Little Rock,
Little Rock, Arkansas, USA
{rxerol, rrejeleene, rbyoung, txmarcoux, mnhussain, nxagarwal}@ualr.edu

***Abstract*—Every minute more than five-hundred hours of video content is uploaded to YouTube, and we can only expect this number to increase. Although YouTube is the most popular video sharing website, studies conducted on this platform are sparse. The lack of effective video analysis techniques presents a tedious challenge for researchers and has hindered overall research on this platform. Due to this, research conducted on YouTube primarily focuses on analyzing text-based content or video metadata. With recent advancements in the development of *Moviebarcode*, a technique that shrinks a movie or video into a barcode, we have developed a tool designed to extend the capabilities of Moviebarcode as a forensic technique for systematically categorizing YouTube videos. We use moviebarcode to summarize an entire YouTube video into a single image to help users understand a video without even watching it.**

***Index Terms*—Moviebarcode, Video Categorization, YouTube, Social Computing Tool**

I. Introduction

In recent years, social media has become ubiquitous among the lives of people who seek to consume content from social media. With respect to content creation and data analysis, it is fair to compare the promises of social media to a modern-day gold rush. Although there are numerous platforms classified as social media; videos have been proven to be the most popular medium for sharing content among users. The most popular platform for video-based content is YouTube.

For every minute, more than five-hundred hours of video is being uploaded to YouTube. We can only expect that number to grow as YouTube focuses on expanding its global reach and making the platform more profitable for content creators [1]. As digital content and consumption is increasing at an incredible rate all over the globe, YouTube video processing becomes computationally intensive. Prior to 2010, YouTube videos could not exceed a video length of 10 minutes. When this restriction was removed, a user published a single video with over 600 hours, which would take 24 days to watch the video [2].

The length of a video is the major limitation for available video processing tools such as computer vision and deep learning based algorithms as they require extensive computational power, time and human effort. In addition to cost and power requirements, currently available video processing tools have a steep learning curve for social computing researchers. Moreover, these tools do not provide analyses to identify videos that are similar or part of same cyber information campaign. Due to these limitations, we extend Moviebarcode, a state of the art video summarization tool that provides linear or close-to-linear processing time regardless of video length.

Moviebarcode is a technique that uses color theory to summarize videos by compressing an entire video into a single image [3]. The result of this technique is a single barcode consisting of generated colors for every frame of the movie. Moviebarcode shows the transitions within videos, gives an overall idea of the video, and enables comparison with other videos without watching the video, thereby saving time.

In this paper, we extend previously described Moviebarcode into an implementation and prototype as a tool to identify similarities among videos, capturing the visual patterns in a video and extract insightful knowledge efficiently. In addition to implementation and prototyping, our novel idea is categorizing videos and identifying botnets and cyber activities with Moviebarcode. We selected YouTube as the first platform to implement our tool. With Moviebarcode, researchers can interact with YouTube video without watching an entire video through summarization. A user is able to optimize important resources such as time to condense each video. For this purpose, we use six different video collections, namely APAC, BalticOps, FifaUnder17Games, ManuGinobiliGames, SpongeBobSquarePants, and HBOSiliconValleyTrailer. The details of the dataset and analysis can be found in Section 4.

The rest of the paper is organized as follows: In Section 2, we describe related works of Moviebarcode. In Section 3, we explain Moviebarcode, its generation process and representation. We describe our dataset, categorization of videos using moviebarcode, discuss our findings and their significance in Section 4. We conclude with major contributions and future direction for this research in section 5.

II. Related Work

Moviebarcode was made popular by Charlie Clark [3], a Tumblr blogger, who generated moviebarcode for numerous movies, and each movie could be filtered by title, director, genre, year. Blogger would capture color patterns in a movie to summarize it, irrespective of its length, to a single barcode.

There are several researchers that used Moviebarcodes for visual video analysis [4, 5]. Manuel Burghardt et al. [4] present an approach that can automatically extract and analyze the language and color parameters from movies by visualizing the most frequent colors of movies. In their approach to visualization, they used clustering algorithms and moviebarcodes. However, we find that there is no summarization tool provided by this research, and their idea falls short of searching and comparing multiple videos. Another study [6] introduced a

Fig. 1. A moviebarcode illustration from a basketball game video.

Fig. 2. A moviebarcode illustration from a soccer game video.

pictorial summary that summarizes a segment of a video for visual representations. Otto et al. [7] presented Moviebarcodes and Long Exposure Images to visualize the colours present in a movie by calculation color population in a frame and stack them together in a moviebarcode format. However, they also normalize the color values to 100. But, their computational cost is more expensive and their research does not include categorization. However, our work is different from aforementioned works as we use moviebarcode for categorizing videos and identifying botnets and cyber activities.

III. Moviebarcode

In this section, we describe Moviebarcode, its generation process and its representation as vector, matrix, tensor. We also explain step by step process used to generate moviebarcodes for videos on YouTube.

A. Moviebarcode

Moviebarcode is a technique to represent a video or a movie as an image by stacking mean values of each frame. Video is a sequence of frames, and there are approximately 30 to 60 frames in each second of a video. When the video is longer than 10 minutes, the number of frames in a video will be greater than 18,000 frames which makes the video analysis even harder because of the high computation requirements. However, Moviebarcode can easily handle any video for analysis.

Moviebarcode is unique to each video. For instance, when the same scene is recorded with the same camera two different times, the Moviebarcode will be different from each other. Furthermore, if a scene is recorded from two different angles, Moviebarcode will again be different. So, it can be said that Moviebarcode is a good technique to catch replicated videos or short clips within a video.

Moviebarcode give dominant colors in each frame. From these dominant colors, significant information about a video can be learned without watching it. For instance, there are two videos; one from a basketball court, and the other from a soccer video. Fig. 1 and 2 show the Moviebarcode of a basketball game and the soccer game videos, respectively, and both Moviebarcode are easily distinguishable. This is important because getting an idea about a video requires significant time to watch and categorize. Moviebarcode technique eliminates this process and shortens the time required for categorizing and filtering videos without watching them.

B. Moviebarcode generation and structure

A moviebarcode is generated from any video or movie, not just limited to YouTube. Since a video has a sequence of frames, each frame is extracted from a video. Then, the mean value of Red (R), Green (G), Blue (B) channels for each frame is calculated. So, after getting a mean value of a frame, a vector of three color values (RGB) is generated (Fig. 3a). By using RGB channels, the gray scale image can also be generated if needed so that the moviebarcode can be represented with a gray scale and used for quantitative analysis. After all these vectors are stacked, we get a matrix of RGB values. So, we can represent a video as a matrix of RGB values (Fig. 3b).

Besides vector and matrix representations, moviebarcodes can also be shown as a tensor. As seen in Fig. 4, when the RGB matrix is converted to tensor, it can be displayed as an image which means representing a video as an image. The width of the image is equal to the number of frames, and the length is to the number of pixels.

The most important question to ask here is what kind of information can be extracted from a moviebarcode. Moviebarcodes use color theory to represent a video. Dominant colors of each frame are stacked on a moviebarcode which means that a moviebarcode keeps dominant colors of the video that are easily identifiable. This sequence of dominant colors and their transitions give information about the video such as changes in the scene, the subject, the narratives of the video within time without watching the video.

C. Generating moviebarcodes from YouTube

Due to YouTube's policy, we do not save the original videos. However, moviebarcode of a video is generated through the process of an algorithm shown in Fig. 5. If a user enters a YouTube video URL, the algorithm firstly checks the availability of this video. If the video is alive and still public to download, we stream the video and generate a moviebarcode on the fly which means that the video is not saved locally. If the URL is for a playlist, the same steps are applied. The algorithm only saves mean values of each frame of a video as .json file.

IV. Datasets and Categorization

For this study, categorization of videos using moviebarcode, we carefully curated a dataset of six different collections of videos. These collections of videos are "APAC", "BalticOps",

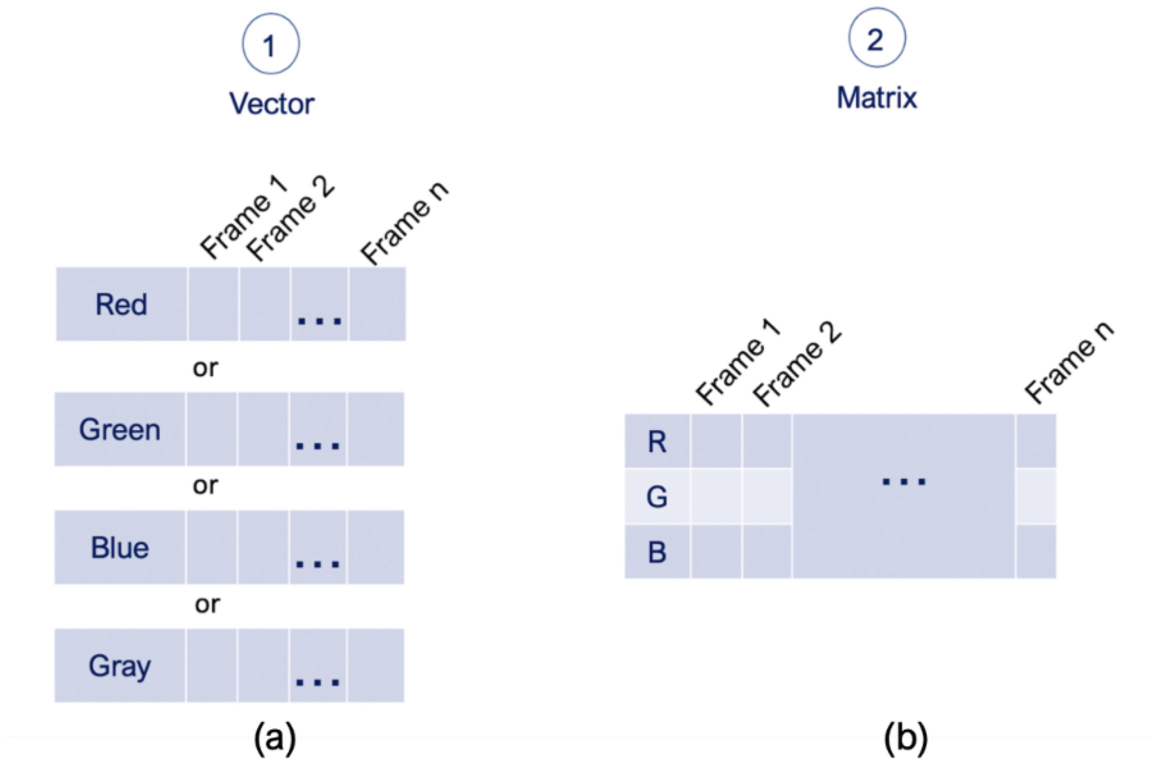

Fig. 3. Moviebarcode representation as a vector or a matrix.

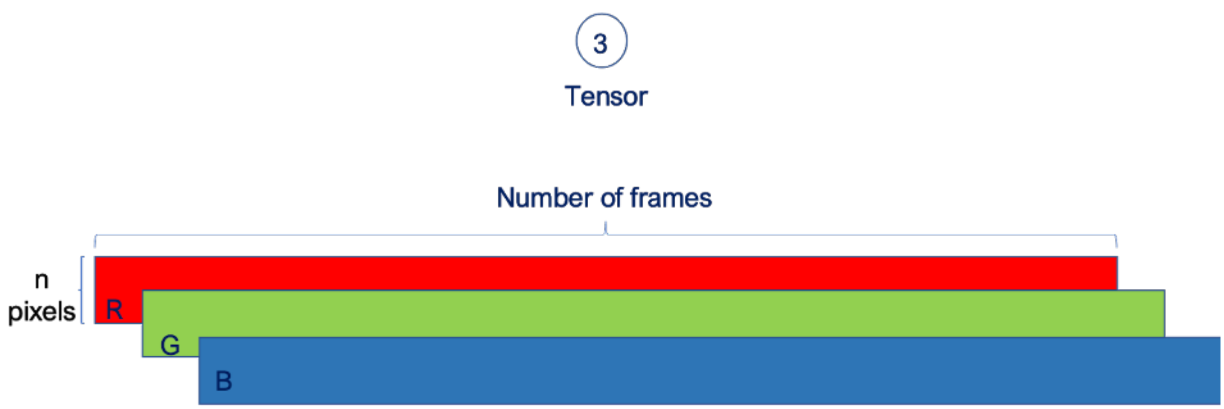

Fig. 4. Moviebarcode representation as a tensor.

"FifaUnder17Games", "ManuGinobiliGames", "HBOSiliconValleyTrailers", and "SpongeBobSquarePants". APAC collection consists of videos with misinformation disseminated about the Asia Pacific region. BalticOps collection consistes of videos with misinformation about NATO's 2019 BALTOPS exercise. FifaUnder17Games collection consists of videos about soccer. ManuGinobiliGames collection consists of videos about the NBA player Manu Ginobili. HBOSiliconValleyTrailers collection consists of trailers of a hit television show called Silicon Valley. SpongeBobSquarePants collection comprises of videos of the cartoon show called Sponge Bob Square Pants. The number of videos in each collection is shown in Table 1.

TABLE I
VIDEO COLLECTION DATASET INFORMATION

Collection Name	Number of Videos
APAC	14
BalticOps	14
FifaUnder17Games	15
ManuGinobiliGames	15
HBOSiliconValleyTrailers	15
SpongeBobSquarePants	15

To construct moviebarcode images, we used tensor

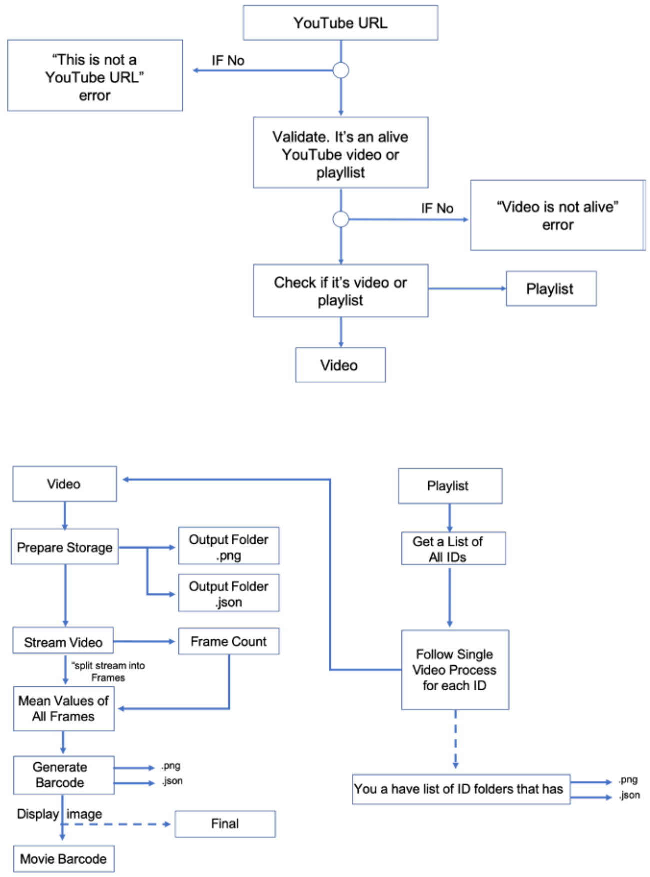

Fig. 5. Algorithm used to generate moviebarcode.

TABLE II
MOVIEBARCODE VIDEO CATEGORIZATION RESULTS

	Precision	Recall	F1-Score	Accuracy
Red channel only	0.79	0.64	0.59	0.64
Green channel only	0.82	0.71	0.69	0.71
Blue channel only	**0.83**	**0.75**	**0.73**	**0.75**
Gray channel only	0.82	0.71	0.69	0.71
All channels together	0.8	0.68	0.64	0.68

Fig. 6. A sample movidebarcode image of a video from HBOSiliconValley collection.

representation. Moviebarcodes have three channels, RGB, and different widths. Each moviebarcode width is equal to the number of frames.

The video categorization pipeline consists of these steps: (1) feature extraction and dimensionality reduction, (2) clustering algorithm, and (3) categorization results. Since our moviebarcode images are not natural images like mentioned in ImageNet dataset [8], we cannot use feature extraction algorithms designed for natural images. We also tried using many different pre-trained convolutional neural network models to extract features with only convolutional layers. However, the result matrix was sparse and did not give us good results on video categorization. Instead, we decided to use moviebarcode pixel values directly on the clustering part of the pipeline.

Next, due to high dimension of images, we applied Principal Component Analysis (PCA) dimensionality reduction algorithm [9]. The results of this step were also used during the clustering step. For the clustering step, we utilized K-means clustering algorithm [10]. The clustering

Fig. 7. A sample moviebarcode image of a video from SpongeBobSquarePants collection.

results are analyzed and compared using the collection labels of videos.

This pipeline was repeated for R, G, B channels and all channels together as well as gray scale. The results of moviebarcode video categorization are shown in Table 2.

Table 2 shows that red channel is not good at distinguishing moviebarcode features from one another. On the contrary, blue channel has the highest scores on performance analysis. It is clearly seen that all channels experiment scores are between red and green channels.

Fig. 6 shows the moviebarcode of a video from the HBOSiliconValley collection. Similarly, Fig. 7 shows the moviebarcode of a video from the SpongeBobSquarePants collection. These moviebarcodes show that it is simple to distinguish one collection from another. Also, it can be clearly seen that scene changes and patterns of similar frames are obvious to catch from moviebarcodes.

Moviebarcodes are useful images that can be used for information retrieval applications such as filtering or grouping images based on their color population. Additionally, the number of different colors in a moviebarcode image can be a good indicator of the pace of the video. For instance, Fig. 6 shows that this video was recorded in the same environment with different camera angles. Same colors are repeated in a sequence and the colors pattern of the video is very similar to other videos in the same collection.

V. Conclusions and Future Work

In this paper, we introduced the use of moviebarcode for video categorization and summarization and showed that video processing is easier with moviebarcode for social computing researchers, especially if they deal with YouTube which is the most popular video sharing platform. Traditional techniques for video categorization are resource intensive and time consuming. Moviebarcode is a great methodology to extract insightful features to describe narratives in a video, capturing visual patterns in a video without watching, and grouping/categorizing similar/same videos together in fast and efficient manner. Results show that using individual channels of a moviebarcode image helps video categorization by differentiating one video from another or grouping them. Each channel carries different features about an image. The idea of splitting channels of an image increased the performance of video categorization.

Moviebarcode technique can be used for further analysis of videos. With the acceleration of new deep learning techniques, it is easy to generate new videos artificially. To identify these artificial videos, moviebarcode can be a great tool to identify similar or same videos, as well as pieces of these videos as a short clip.

RGB channels are used in this study, but YCbCr or HSV color channels can also be used to categorize videos. Each color channel has different features about a video. Color theory techniques show that different color channels can be used for different purposes. With this motivation, video categorization can be examined by using other color channels different from RGB. Also, other data models such as transcription of a video or metadata can be combined for video categorization. These multiple data models may boost performance of video categorization.

Acknowledgment

This research is funded in part by the U.S. National Science Foundation (IIS-1636933, ACI-1429160, and IIS-1110868), U.S. Office of Naval Research (N00014-10-1-0091, N00014-14-1-0489, N00014-15-P-1187, N00014-16-1- 2016, N00014-16-1-2412, N00014-17-1-2605, N00014-17- 1-2675), U.S. Air Force Research Lab, U.S. Army Research Office (W911NF-16-1-0189), U.S. Defense Advanced Research Projects Agency (W31P4Q-17-C-0059) and the Jerry L. Maulden/Entergy Fund at the University of Arkansas at Little Rock. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding organizations.

Nomenclature

RGB: Red, Green, Blue
YCbCr: Luma, Blue-difference chroma, Red-difference chroma components
HSV: Hue, Saturation, Value
PCA: Principal Component Analysis

References

[1] P. Suciu. "Is It Possible To Become The Next Big YouTube Star In 2020?," *Forbes*, 03-Jan-2020.

[2] Y. Press. "Up, Up and Away - Long videos for more users." *YouTube*. [Online]. Available: https://youtube.googleblog.com/2010/12/up-up-and-away-long-videos-for-more.html. [Accessed: 16-Mar-2020].

[3] "Moviebarcode," *Tumblr*. [Online]. Available: https://moviebarcode.tumblr.com/. [Accessed: 16-Mar-2020].

[4] M. Burghardt, M. Kao, and C. Wolff. "Beyond Shot Lengths – Using Language Data and Color Information as Additional Parameters for Quantitative Movie Analysis." *In Digital Humanities 2016: Conference Abstracts*. Jagiellonian University and Pedagogical University, Kraków, pp. 753-755, 2016.

[5] M. Barbieri, G. Mekenkamp, M. Ceccarelli, and J. Nesvadba. "The color browser: a content driven linear video browsing tool," *IEEE International Conference on Multimedia and Expo, (ICME 2001)*, Tokyo, Japan, pp. 627-630, 2001.

[6] M. M. Yeung and Y. Boon-Lock. "Video visualization for compact presentation and fast browsing of pictorial content." *IEEE Trans. Circuits Syst. Video Techn*, vol 7, pp. 771-785, 1997.

[7] I. Otto, A. Plutino, M. Lanaro, and A. Rizzi. "All the colours of a film: A study on the chromatic variation of movies." *AIC Interim Meeting*, Lisbon, Portugal, 2018.

[8] J. Deng, W. Dong, R. Socher, L.J. Li, K. Li and L. Fei-Fei. "ImageNet: A Large-Scale Hierarchical Image Database." *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009.

[9] Scikit Learn. "Decomposing signals in components (matrix factorization problems)." *Scikit Learn*. [Online]. Available: https://scikit-learn.org/stable/modules/decomposition.html#pca. [Accessed: 22-Mar-2020].

[10] Scikit Learn. "K-means." *Scikit Learn*. [Online]. Available: https://scikit-learn.org/stable/modules/clustering.html#k-means. [Accessed: 22-Mar-2020].