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 and later use cluster them based on similarity. We analyzed six video collections and using moviebarcode only and without looking at the video content, we were able to achieve an accuracy of 75%. Using our method, an analyst can quickly group videos into bin computationally reducing the overhead of manually doing 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].

There are many available deep learning based video categorization studies [3, 4]. These studies show great contribution to the research community. However, 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 directly provide information to use in identification of cyber activities on videos. 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 [5]. The result of this technique is a single barcode consisting of generated colors for every frame of the movie. Moviebarcode shows the color transitions within videos, gives an overall idea about the video content, 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 with moviebarcode. For this purpose, we created six different video collections, namely APAC, BalticOps, FifaUnder17Games, ManuGinobiliGames, SpongeBobSquarePants, and HBOSiliconValleyTrailer. Using categorization algorithm to group the moviebarcodes and got promising results that are explained under section 4. 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. 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 Clark [5], 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 [6] such as ColorBrowser [7]. Burghardt

 ISBN: 978-1-61208-800-6

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

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

et al. [6] present an approach that can automatically extract and analyze the language and color parameters from movies by visualizing the most frequent colors in movies. In their approach to visualize, 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 [8] introduced a pictorial summary that summarizes a segment of a video for visual representations. Otto et al. [9] 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.

Our work is different from aforementioned works as we use moviebarcode for categorizing videos. Also, the method to group videos based on similarity has been done empirically and requires an analyst to manually watch the videos to group them. Our method reduces the effort significantly by using computational methods.

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 gives 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 can be generated for 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. 3.1). 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. 3.2).

Besides vector and matrix representations, moviebarcodes can also be shown as a tensor. As seen in Fig. 3.3, 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 that a user can assign. This number of pixels is 224 in our experiments.

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 can 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

For data collection and streaming, we use public data API of YouTube [10] to download data from YouTube, and

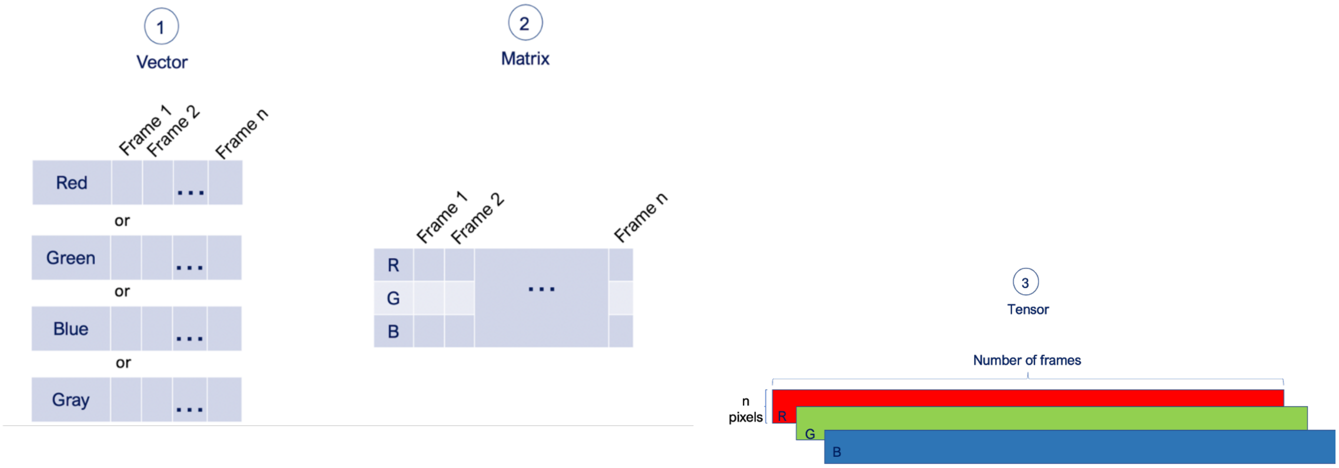

Fig. 3. Moviebarcode representations as a vector (1), matrix (2) and tensor (3).

OpenCV computer vision framework [11] to stream videos from YouTube. However, due to YouTube's policy, we do not save the original videos. Moviebarcode of a video is generated through a process shown in Fig. 4. If a user enters a YouTube video URL, the procedure first checks the availability of this video. If the video is online 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 recursively for each video. 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. Subject Matter Experts (SME) helped us identify and group the videos that were later collected using data collection method described in our previous studies [12,13]. These collections of videos are "APAC", "BalticOps", "FifaUnder17Games", "ManuGinobiliGames", "HBOSiliconValleyTrailers", and "SpongeBobSquarePants". APAC collection consists of conspiracy theories and misinformation videos being disseminated related to various events and issues in the Asia Pacific region. BalticOps collection consists of videos with misinformation about NATO's 2019 BALTOPS exercise. FifaUnder17Games collection consists of videos about soccer. ManuGinobiliGames collection consists of videos about highlights from the NBA games that Manu Ginobili plays. HBOSiliconValleyTrailers collection consists of trailers of a hit television series 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. The lengths of videos in video collections ranges from 3 minutes to 20 minutes.

To construct moviebarcode images, we used matrix representation. Moviebarcodes have three channels, RGB, and different widths. Each moviebarcode's width is equal to the number of frames.

TABLE I
VIDEO COLLECTION DATASET INFORMATION

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

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

The video categorization pipeline consists of these steps: (1) image pre-processing to align all moviebarcodes to the same shape in terms of width and length, (2) applying dimensionality reduction algorithm to all input datasets, (3) applying a clustering algorithm to group similar moviebarcodes into the same clusters, and (4) comparing cluster results with the video collection labels of videos for evaluation. The performance of the process is measured with confusion matrix [14]. We 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. Since our moviebarcode images are not natural images like ImageNet dataset [15], moviebarcodes require custom feature extraction algorithm. Instead, we decided to use one of the most important features of an image which is pixel value directly on the clustering part of the pipeline. The fine tuning of convolutional neural networks for better feature extraction and alternative video

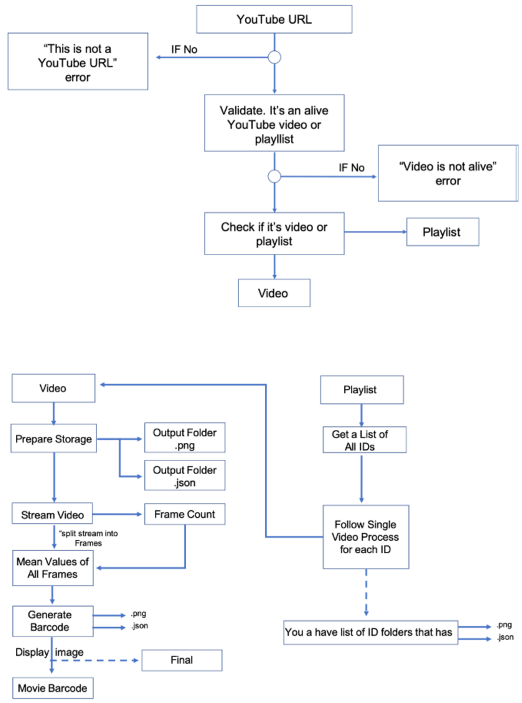

Fig. 4. Algorithm used to generate moviebarcode.

Fig. 5. A sample movidebarcode image of a video from HBOSiliconValley collection.

Fig. 6. A sample moviebarcode image of a video from SpongeBob-SquarePants collection.

categorization with a moviebarcode dataset is left for future studies.

Next, due to high dimension of images, we applied Principal Component Analysis (PCA) dimensionality reduction algorithm [16]. The results of this step were used during the clustering step. Due to its simple nature to implement and run, we utilized K-means clustering algorithm [17] with the cluster value as the number video collections for the clustering step. Next, we applied model evaluation with confusion matrix. We repeated the process of k-means and model evaluation 10 times and calculated the average of these experiments for the final result. The clustering results are analyzed and compared using the collection labels of videos.

This pipeline was applied on tensor of moviebarcodes which are all color channels together. After that, we repeated the process for individual channels and gray scale. The all results of moviebarcode video categorization are shown in Table 2.

Table 2 shows that red channel in moviebarcode is not a good feature to distinguish the clusters. On the contrary, blue channel has the highest scores on all metrics including precision, recall, f1-score, and accuracy. The scores for all other channels and their combinations are between red and blue channels.

Fig. 5 shows the moviebarcode of a video from the HBOSiliconValley collection. And in contrast, Fig. 6 shows the moviebarcode of a video from the SpongeBobSquarePants collection. These moviebarcodes show that it is simple to distinguish one collection from another. Also, changes in the scenes and patterns of similar frames can be clearly observed 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.

 ISBN: 978-1-61208-800-6

V. Conclusions and Future Work

In this paper, we introduced the use of moviebarcode for video categorization and summarization. We also demonstrated that the video processing is easier with moviebarcode for social computing researchers, especially if they deal with YouTube which is the most popular video sharing platform. Our experiments focus on reducing the video to colors. Traditional techniques for video categorization are resource intensive and time consuming. Moviebarcode is a great methodology to extract insightful features by capturing visual patterns in a video without watching, and grouping or categorizing same or similar videos together in fast and efficient manner.

Results show that using individual channels of moviebarcode image helps video categorization by differentiating one video from another or grouping them. Each channel carries different features about an image. Splitting the channels of an image increased the performance of video categorization. Our findings suggests that analyzing only the colors within the video without looking the video content in detail gives the accuracy of 75%.

Video length is one of the most important features about the video. But, it is also one of the limitations of moviebarcode technique because it is difficult to align long videos with short videos. In this paper, we experimented with six video collections. In order to make our model more generalized, future research could examine the experiment pipeline of our model on other video collections. Since we use moviebarcode pixels directly on the categorization pipeline, it would be better to have a custom feature extraction method to extract more features from the moviebarcodes.

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 artificially generated videos, moviebarcode might be a great tool to identify similar or same videos, as well as pieces of these videos as a short clip. Even though we currently use moviebarcod only video categorization, we could use them to detect scene changes and narratives by detecting changes in colors.

RGB channels are used in this study, but YCbCr or HSV color channels could 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 could be examined by using other color channels different from RGB. Other data models such as transcription of a video or metadata could also be combined for video categorization. These multiple data models might 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] Y. Jiang, Z. Wu, J. Wang, X. Xue and S. Chang, ”Exploiting Feature and Class Relationships in Video Categorization with Regularized Deep Neural Networks,” *in IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 2, pp. 352-364, 1 Feb. 2018, doi: 10.1109/TPAMI.2017.2670560.

[4] M. Liu, L. Nie, X. Wang, Q. Tian and B. Chen, ”Online Data Organizer: Micro-Video Categorization by Structure-Guided Multimodal Dictionary Learning.” *in IEEE Transactions on Image Processing*, vol. 28, no. 3, pp. 1235-1247, March 2019, doi: 10.1109/TIP.2018.2875363.

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

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] YouTube Data API. ”Add YouTube functionality to your app” *YouTube*. [Online]. Available: https://developers.google.com/youtube/v3/

[11] *OpenCV*. [Online]. Available: https://opencv.org/

[12] J. Kready, S. A. Shimray, M. N. Hussain, and N. Agarwal. ”YouTube data collection using parallel processing.” *In 2020 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)*. (pp. 1119-1122), IEEE, May 2020.

[13] M. N. Hussain, S. Tokdemir, N. Agarwal, and S. Al-Khateeb. ”Analyzing disinformation and crowd manipulation tactics on YouTube.” *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE, 2018.

[14] S. Visa, B. Ramsay, A. Ralescu, and E. van der Knaap. ”Confusion matrix-based feature selection.” *MAICS*. 710, pp.120-127, 2011.

[15] 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.

[16] 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].

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