Scene Change Detection using Different Color Pallets and Performance Comparison

F. Bulut, and S. Osmani

Abstract— **In the world of massive uploaded videos, to be able to cover the content of a video at a glance becomes a necessity since there is no enough time to watch the whole video for an individual. Looking at frames of different scenes in a video gives a brief idea of the content, when each different scene images are listed to be checked by the user. In this study, an approach using various color palettes is proposed to be able to detect the different scenes of a video. In the proposed method, color histogram values of sequential frames firstly are calculated. If the difference in the histogram values of the pair frames in sequence is over a threshold value (percentage of change), scene change is detected. In the experimental studies, 3-Bit RGB (Red Green Blue), 6-Bit RGB, 8-Bit RGB, 9-Bit RGB, 1-Bit Binary, 4-Bit Gray, and 8-Bit Gray palettes are implemented over a list of video files and compared. In the comparisons of palettes, accuracy, precision, recall, and F1-Score performance metrics are used. In the performance accuracy controls, 6-Bit RGB color pallet with a threshold level value of 35% has been experimented as the best of all.**

Index Terms— **Scene change detection, RGB and gray palettes, histogram.**

I. Introduction

IN the recent world of technology, electronic devices and social media are able to store huge sized video files where it makes searching and finding the required videos harder and time consuming. In this case, understanding the content of a video just by looking at its name is not considered as a correct and efficient way. Also, the meta information of a video might be irrelevant with the real content of the video. However, tagging method has been initially released as a sole solution. But this method brings out two more critical issues. First, if the irrelevant tag names were given mistakenly or knowingly, then the whole search process would be negatively affected. Second, it can be more time consuming.

Previously researchers have concluded static and dynamic video summarizations under the same field while stationary image summarizations provided a small collection of photos achieved from video sources.

Dynamic video summarizations have constructed from relatively queued photos and original video's audio. These two summarization techniques are very different from each other, stationary image summarizations dealt only with images no need for audio, where dynamic image summarizations dealt with images along with audio which made it to worked very slowly.

As it is known video files are made from consecutive static images, named as frames. One second of a video file contains a certain amount of sequential frames, called as frame per second (FPS). The primitive TV broadcasting system releases 24 FPS [1] whereas this value is 30 FPS, in so many other video formats. In the HDTV it is 50 FPS. The value of FPS with interpolation techniques in HEVC (High Efficiency Video Coding) types of industries is rising up to 300 and with the resolution 8192x4320 pixels [2].

Scene change detection using histogram and color palettes is an answer to so many questions and challenges which are faced in world's many different sectors such as education sector, filming sector and most importantly security sector.

Internet offers free learning resources but its hectic to find the desired video tutorials in seconds. For finding the desired video the user is supposed to keep on searching and playing each and every video in order to get sure of its content not only this method is time consuming but sometimes it ends up without a positive result.

While scene change detection using histograms and color pallets allows users to find the desired video in lesser than a minute, as the user entered the required information the system gives as output the most important frames of the video, so the user can check the content and decide.

Also, scene change detection with histogram and color pallets brings an ease in security sector. As an example, let's suppose a 24-hour static camera in front of a grocery shop, in case if a thief enters, the camera records it. In a similar case finding out the true period of the theft is time consuming.

In contrast the other scene change detection techniques, here the proposed system allows the users to input their recorded file, provides the required information as an output the system and gives the most important frames. Therefore, the information about the incident can be found easily.

Meanwhile, by using scene change detection with histograms and color pallets algorithm many intelligent systems can be built. For example, an intelligent camera with static position, where if the camera changes its position by any outside factor it should be able to compare the initial position's information with the current and turn back to its initial position.

This paper contains four more sections. In the second one, there is a survey including the related studies. In the third section, the details of suggested method will be explained.

✉ **F. BULUT**, Istanbul Rumeli University, Computer Engineering Department, Istanbul, TURKEY, (e-mail: faruk.bulut@rumeli.edu.tr).

S. OSMANI, Data Analyst, Netlinks Technology, Kabul, AFGHANISTAN, (e-mail: shaira@netlinks.ws).

Manuscript received May 16, 2017; accepted August 11, 2017.
DOI: 10.17694/bajece.336217

Before the conclusion section, there are some experimental results and performance comparisons of the proposed methods

II. RELATED STUDIES

Since video summarization is one of the big questions in this field, some previous researches have already proposed some methods. Some valuable methods are selected and listed chronologically below.

In 1996, Wolf released a solution for video summarization technique, identifying the main frame shots from the video [3]. It is preferred to use optical flow computations in order to identify local minimum of motion in a shot-stillness. This technique allows to discover both gestures and camera motions. Results of the effectively summarized shots have showed that this algorithm is successfully select many key frames from a single complex shot.

Defaux in 1996 explained an algorithm for summarization in which the image is identified in terms of the primitives of the scene [4]. In this technique representing of the video dataset in fewer bits, compression of files plays a great role. The proposed system utilizes a specific decoder to interpolate the frames in order to provide a reliable system.

Hong Zhang and his friends in 1997 released a solution for video summarization techniques as its system provided methods for temporal segmentation of video sequences into individual camera shots, using a novel twin comparison method [5]. This method was capable of detecting both camera shots implemented by sharp break and gradual transitions implemented by special editing techniques, including dissolve, wipe, fade-in and fade-out; and content based key frame selection of individual shots by examining their temporal variation of video contents and selecting a key frame after the differences of contents between the current frame and a preceding selected key frame exceeds a set of preselected thresholds.

Huang and his friends in 1999 released another solution in this filed. They proposed a new technique that aggregates motion information and intensity in order to discover scene changes in both sudden scene changes and gradual scene changes. Two major attributes are taken as the basic dissimilarity measures, and self Additionally, they propose a new intensity statistics model in order to find gradual scene changes. Their experimental studies proved that the proposed method had outperformed the previous approaches up to 1999.

The other researchers, Lee and his friends in 2003 explained that video summaries allow condensed representations of a video content through a combination of continuous images, graphical representations, video segments, and textual descriptors [7]. Their framework distinguishes between video summaries and video summarization techniques.

In 2014, in large video collections with clusters of categories, Daniela Potapov wrote an article using machine learning techniques. He explained that category-specific video summarization can produce more accurate video summaries than unsupervised learning methods that are blind to the video categories [7]. Their approach firstly performed a temporal segmentation into semantically-consistent segments providing a video from a known category. Then, equipped with a Support Vector Machine classifier, this approach assigned some scores to each of the segments. The resulting video assembles the segment sequences with the bigger scores. Therefore, the gained short video summary is highly informative. Experimental studies over a list of video files proves that the technique gives relevant video summaries.

Zhang and his friend in 2015 released another technique as Multi-video summarization. In this research, the proposal is a new summarization method that implements preserves well-aesthetic frames and video stability. Particularly, a multi-task attribute selection is used in order to efficiently detect the semantically important attributes. Then, the key frames are selected based on their contributions in order to rebuild the video semantics. Then, a probabilistic model is suggested to fit the key frames dynamically into the video summary [9].

Lately in 2017, Wu and his friends has newly proposed a scene change detection method especially focused on between multi-temporal image scenes. In their research, a novel scene change detection method via Kernel Slow Feature Analysis (KSFA) is proposed. With the help of post-classification fusion, KSFA is used to extract the nonlinear temporally invariant attributes for better measure between corresponding multi-temporal image scenes. The post-classification fusion methods are based on Bayesian theory. They have experimented that the proposed method increases the accuracy of scene change detection, scene transition identification, and scene classification.

In this scene change detection field, different methods with their variations are proposed in the literature in the last decades. Most of the academic publications in this area stands on the empirical studies engineering field whereas few of the studies include novel approaches.

III. PROPOSED METHOD

Experimental study is performed on a list of selected videos. Collected video files which have different types and specifications are listed in Table 1. In the list, there are four video files whose attributes are as resolution, type, color type, aspect ratio (ration in dimensions), FPS (Frame Per Second), number of total frames, number of total different scene, duration in seconds, and lastly the source which comes from. As it is seen, these collected videos which have discrete features crate a rich experimental area where different types of pallets can be compared. Each video file has different frame per second rate.

In the RGB true color system, the capacity of each of color is 8 bits. The combination of the colors represents the picture

Additionally, in Table 1, the total number of different scenes in each video can be seen at the last column. The list has been prepared by a human in order to check the performance of the implemented methods in the experiments.

All the experiments are performed in the MATLAB environment [11] on MACOS operating system. The process firstly starts step by step with the decomposition of the frames in each video files.

The original resolutions, color system, and the FPS rates of the video files in the test list are bigger than normal for a digital image video processing operation.

TABLE I
FEATURES OF VIDEO FILES

Video ID	Resolution	Category	Color Type	Aspect Ratio	FPS	Duration	# of total frames	# of total different scene
1	640×480	VGA	24-bits RGB	4:3	24	32 sec.	24×30=720	5
2	1280×720	HD Ready	24-bits RGB	16:9	25	60 sec.	25×60=1508	9
3	1920×1080	Full HD	24-bits RGB	16:9	30	125 sec.	30×120=3602	17
4	3840×2160	4K	24-bits RGB	16:9	50	58 sec.	50×50=2500	11
				Total	275 sec.	8330		

Those given numerical values in the list increase both the time complexity and the computational complexity. Hence, both the size of the videos and the color bits should be decreased for a fast execution. In the empirical studies, some quantization techniques are implemented. As it is known, quantization can be called as summarization of videos.

In Table 2, there are 7 different color pallets used for quantization techniques. As it is known, color palettes are the range of colors used in a visual medium. Also, a quantization (summarization) technique is done in bit values of the RGB color values [12]. For example, 8-bit value of a color in RGB transforms into 2-bit value by deleting the rest of first 2 bits. In 8 bit colors, the coming numbers after the initial numbers are the details and tones of the specific color. The most important color codes are accepted as the initial bits in image processing.

The resolutions of the videos and the FPS rates have remained the same throughout the process. It is certain that to reduce those values will also decrease the computational time. Since this study focuses on the color pallets, these things have not changed in the experiments. Probably in further studies, these variations can be tested.

TABLE II
COLOR PALETTES FOR QUANTIZATION TECHNIQUES

Videos ID	Palette name	R	G	B	Gray	Color Bits	Color Status	Order Status
1	3-bit RGB	1	1	1	-	$2^3 = 8$	colored	order
2	6-bit RGB	2	2	2	-	$2^6 = 64$	colored	order
3	8-bit RGB	3	3	2	-	$2^8 = 256$	colored	Not-order
4	9-bit RGB	3	3	3	-	$2^9 = 512$	colored	order
5	1-Bit Binary	-	-	-	1	$2^1 = 2$	black-white	order
6	4-Bit Gray	-	-	-	4	$2^4 = 16$	gray	order
7	8-bit Gray	-	-	-	8	$2^8 = 256$	gray	order

RGB stands for red-green-blue on a computer display. These three basic colors can be combined in various proportions in order to obtain any color in the visible spectrum for a human. In the combination, each levels of red, green and blue can vary from 0 to 100 percent of full intensity. If the total bit capacity is 24 of the RGB type, each level is represented by the range of decimal numbers from 0 to 255 since 28 = 256.

A histogram is a specific statistical information that displays the frequency of color bits in successive numerical intervals of equal size. In common forms of histogram, the dependent variable is shown along the vertical axis and the independent variable is shown along the horizontal axis.

The proposed system captures each frame when processing on a selected video file. In each frame, the color histogram values for Red, Green and Blue are calculated. As an example, red, green, blue and RGB histogram values can be seen using the Lena picture in Fig. 1, Fig. 2, Fig. 3, and Fig. 4 sequentially.

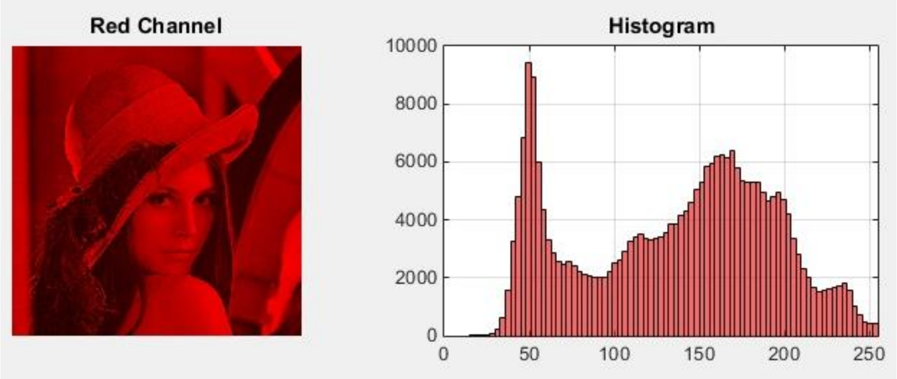

Fig. 1. Red color histogram

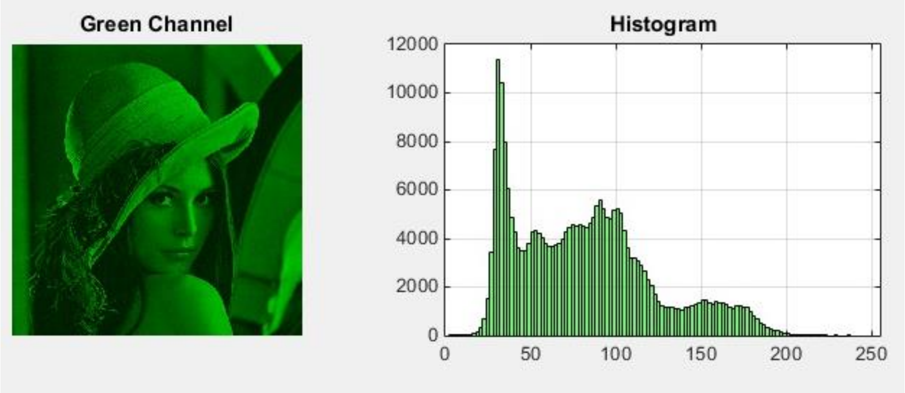

Fig. 2. Green color histogram

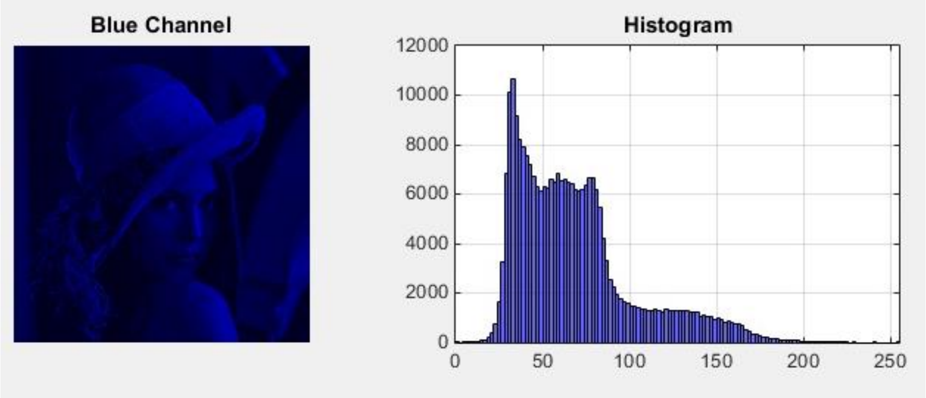

Fig. 3. Blue color histogram

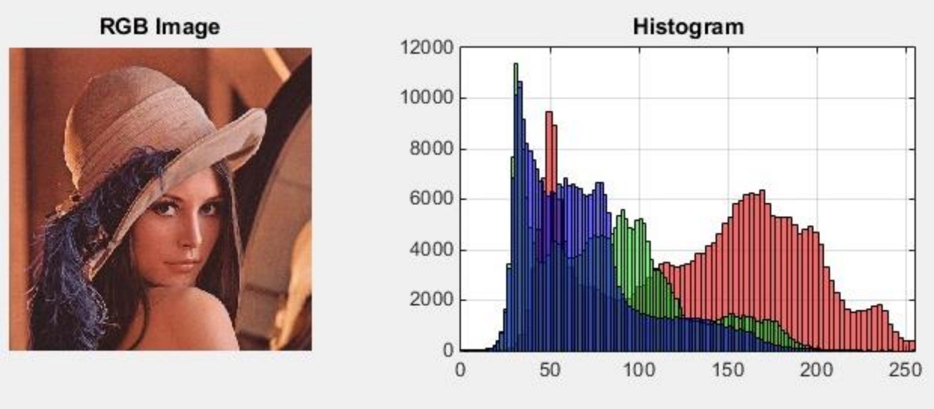

Fig. 4. Red Green and Blue color histograms

Those histogram values of the colors are different from each other. But in this example this values are taken from a static picture, Lena [13]. In video processing, these calculations are done for each frame. It should be emphasized that the histogram values continuously change in each frame during the video.

Basically the algorithm focuses on the changes of color histogram values on sequential frames. While the system generates histogram chart for each captured frame, the changes should be checked in a sequential pair of frames in order to detect the scene changes. For example, in a pair of frames, if there is a small change in the Blue histogram, it might be that a small blue colored ball has newly entered the current scene. If there is big change in the histogram rates, that means the previous scene has changed.

How can be the “small” and the “big” changes defined in the histogram values? The answer of the question affects the accuracy and the performance of the proposed system. The answer is the threshold value that compares the differences of histogram frequencies between each continuous frame. If the total histogram change is bigger than the specified threshold value, it means a new scene occurs. If not, it is the same repeating scene.

IV. EXPERIMENTAL RESULTS AND PERFORMANCE COMPARISON

In order to evaluate the accuracy, the system should use some basic performance metrics such as True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). These basic metrics are the basis of other well-known criterions such as accuracy, precision, recall, and confusion matrix table [14].

The proposed software mechanism sometimes detects the real scene change. This is called as TP detection. Sometimes the system detects a scene change when actually it is not. This is called as FP. If the system cannot detect the scene change, it is called as FN. And lastly if the system detects any change when there is no change, it is called as TN. TN value in this application cannot be realized. Because of this, TN will always be 0 throughout the experiments. These performance criterions of TP, TN, FP, and FN here are newly adapted to this study.

The TP, TN, FP and FN values are manually calculated by watching the whole videos in the list. The TP, TN, FP and FN rates are as shown in Table 3.

TABLE III
TP, TN, FP AND FN VALUES OF PALLESTS OVER VIDEOS

	Video 1				**Video 2**				**Video 3**				**Video 4**			
	TP	**TN**	**FP**	**FN**	**TP**	**TN**	**FP**	**FN**	**TP**	**TN**	**FP**	**FN**	**TP**	**TN**	**FP**	**FN**
3-Bit RGB	2	-	4	3	5	-	5	4	10	-	14	7	5	-	4	6
6-Bit RGB	3	-	2	2	7	-	4	2	15	-	9	2	9	-	3	2
8-Bit RGB	5	-	1	0	8	-	1	0	17	-	2	0	11	-	1	0
9-Bit RGB	5	-	1	0	8	-	2	0	17	-	1	0	11	-	0	0
1-Bit Binary	2	-	3	3	4	-	5	6	5	-	11	12	6	-	4	5
4-Bit Gray	3	-	3	2	5	-	4	6	10	-	9	7	8	-	3	3
8-Bit Gray	3	-	4	2	5	-	4	5	11	-	8	6	9	-	3	2

In the experiments, different threshold levels such as 10%, 20%, 30%, 35%, 40%, and 50% have been tested for better performance. 35% level has been experimented as the best of all.

The other performance criterions are accuracy, precision, recall, F1-Score, and recall are as follows.

Accuracy (Acc), more commonly, is a description of systematic errors, a measure of statistical bias. The accuracy formula is in the Eq. (1).

$$Accuracy = \frac{Correct\ predictions}{Total\ number\ of\ data} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Precision, also called as positive predictive value, is the fraction of retrieved instances that are relevant. Precision can be calculated using the formula in the Eq. (2).

$$Presicion = \frac{True\ predicted}{Number\ of\ predictions} = \frac{TP}{TP + FP} \quad (2)$$

Recall, also known as sensitivity, is the fraction of relevant instances that are retrieved. Recall formula can be seen in the Eq. (3).

$$Recall = \frac{Predicted\ value}{Real\ value} = \frac{TP}{TP + FN} \quad (3)$$

F1 Score is the harmonic mean of precision and recall both. F1 gives more accurate results than precision or recall because it contains these metrics. F1-Score formula is in the Eq. (4).

$$F1\ Score = \frac{2}{\frac{1}{P} + \frac{1}{R}} = 2 \times \frac{P \times R}{P + R} \quad (4)$$

TABLE IV
ACCURACY, PRECISION, RECALL AND F1 VALUES OVER FOUR VIDEOS

	Video 1				Video 2				Video 3				Video 4			
	Acc	**P**	**R**	**F1**	**Acc**	**P**	**R**	**F1**	**Acc**	**P**	**R**	**F1**	**Acc**	**P**	**R**	**F1**
3-Bit RGB	0.22	0.33	0.40	0.36	0.36	0.50	0.56	0.53	0.32	0.42	0.59	0.49	0.33	0.56	0.45	0.50
6-Bit RGB	0.43	0.60	0.60	0.60	0.54	0.64	0.78	0.70	0.58	0.63	0.88	0.73	0.64	0.75	0.82	0.78
8-Bit RGB	0.83	0.83	1.00	0.91	0.89	0.89	1.00	0.94	0.89	0.89	1.00	0.94	0.92	0.92	1.00	0.96
9-Bit RGB	0.83	0.83	1.00	0.91	0.80	0.80	1.00	0.89	0.94	0.94	1.00	0.97	1.00	1.00	1.00	1.00
1-Bit Binary	0.25	0.40	0.40	0.40	0.27	0.44	0.40	0.42	0.18	0.31	0.29	0.30	0.40	0.60	0.55	0.57
4-Bit Gray	0.38	0.50	0.60	0.55	0.33	0.56	0.45	0.50	0.38	0.53	0.59	0.56	0.57	0.73	0.73	0.73
8-Bit Gray	0.33	0.43	0.60	0.50	0.36	0.56	0.50	0.53	0.44	0.58	0.65	0.61	0.64	0.75	0.82	0.78

In Table 4, Accuracy and F1 scores are calculated and the results are listed using the corresponding formulas over each video file.

Accuracy and F1 score rates are illustrated in bar charts in Figure 5 and Figure 6 respectively. In both Accuracy and F1 metrics, the performance of each color pallets can be examined at a glance. 8-Bir RGB and 9-Bit RGB color pallets outperform when compared with the others.

1-Bit Binary image is the worst one, of all. The main reason of the failure is the loss of image details. As it is known, the binary image contains just black and white information. There is not any other color. Hence, the details are disappeared in the quantization phase.

The similar situation to the 1-Bit Binary image exists in the gray color pallets. The unsuccessful performance of 4-Bit and 8-Bit gray level pallets can be seen in Figure 5 and Figure 6. According the experimental results increasing the bit number of a gray colored pallet will not rise the accuracy level. Thus, it can be inferred that red, green and blue color values are very important in the scene change detection.

When compared all the color pallets, the 8-Bit and 9-Bit RGB methods give the best performance in detection of scenes. Maybe 8-Bit method can be preferred to 9-Bit one because of less computational and space complexity.

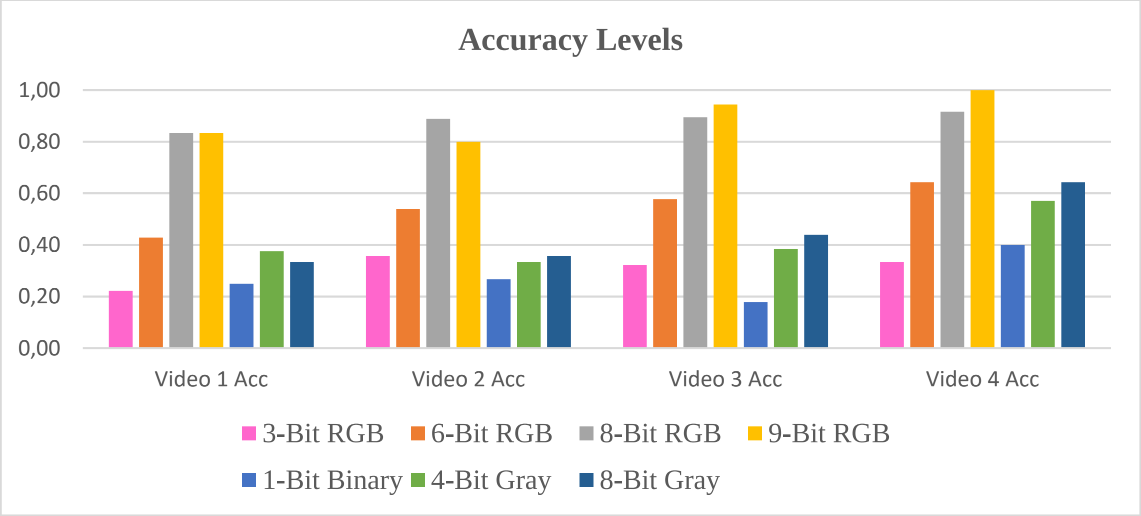

Fig. 5. Accuracy values over videos.

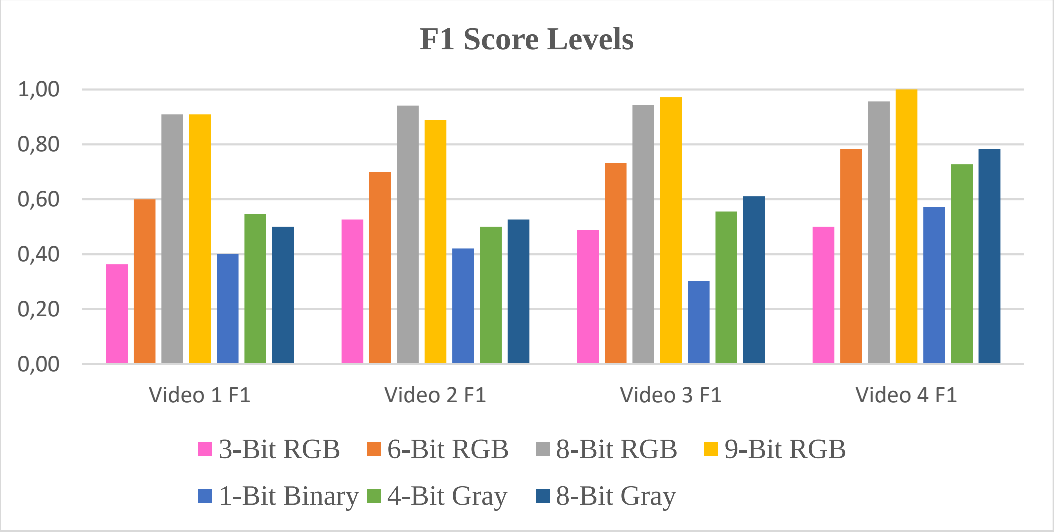

Fig. 6. F1-Score values over videos.

V. Conclusions

In this project we aim to develop a system which will summarize and give content of any lengthy video. This system will provide so many efficiencies for the user its aimed to improve the performance of searching the content of any video with less time consumptions. This proposed mechanism is composed of some techniques in order to do the summarization and give the content as the system uses color reduction, color pallets slow motion detection and the histogram charts. Proposed model of summarization can be concluded as the fact that best scene change detection can be performed using 8-RGB color pallet. The current concern is about saved static videos but in further studies this concern might be to develop a real time system which summarizes online social media videos. Summarization has been an important issue in world of image processing so the research field has been worldwide and many solutions has been implemented.

References

[1] Nakamura, J. (Ed.). (2016). Image sensors and signal processing for digital still cameras. CRC press.

[2] Rao, K. R. (2016, September). High efficiency video coding. In Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA), 2016 (pp. 11-11). IEEE.

[3] Wolf, W., (1996), “Key Frame Selection by Motion Analysis”, Proceeding of IEEE International, Conference on Acoustics, Speech and Signal Processing, Atlanta, GA, 1228-1231

[4] Dufaux, Frédéric, and Fabrice Moscheni. "Segmentation-based motion estimation for second generation video coding techniques." Video Coding. Springer US, 1996. 219-263.

[5] Zhang, Hong J., Jian H. Wu, and Stephen W. Smoliar. (1997), "System for automatic video segmentation and key frame extraction for video sequences having both sharp and gradual transitions." U.S. Patent No. 5,635,982. 3 Jun. 1997.

[6] Huang, C. L., & Liao, B. Y. (2001). A robust scene-change detection method for video segmentation. IEEE Transactions on Circuits and Systems for Video Technology, 11(12), 1281-1288.

[7] Lee, J., Lee, G. ve Kim, W., (2003), “Automatic video summarizing tool using MPEG-7 descriptors for personal video recorder”, IEEE Trans. on Cons. Elect, Vol 49, 49-742

[8] Danila Potapov, Matthijs Douze, Zaid Harchaoui, Cordelia Schmid. Category-specific video summarization. ECCV 2014 - European Conference on Computer Vision, Sep 2014, Zurich, Switzerland, Springer, 2014

[9] Zhang, L., Xia, Y., Mao, K., Ma, H., & Shan, Z. (2015). An effective video summarization framework toward handheld devices. Industrial Electronics, IEEE Transactions on, 62(2), 1309-1316.

[10] Wu, C., Zhang, L., & Du, B. (2017). Kernel slow feature analysis for scene change detection. IEEE Transactions on Geoscience and Remote Sensing, 55(4), 2367-2384.

[11] Marques, Oge. Practical image and video processing using MATLAB. John Wiley & Sons, 2011.

[12] Umbaugh, Scott E. Digital image processing and analysis: human and computer vision applications with CVIPtools. CRC press, 2016.

[13] Lena Söderbergt, Image Processing Benchmark image, 1973. URL: tps://en.wikipedia.org/wiki/Lenna, taken date: 06/02/2017.

[14] Mousavizadegan, M., & Mohabatkar, H. (2016). An Evaluation on Different Machine Learning Algorithms for Classification and Prediction of Antifungal Peptides. Medicinal Chemistry, 12(8), 795-800

Biographies

Faruk BULUT was born in Kayseri, Turkey in 1974. He got his bachelors’ degree in the Computer Education Department at Marmara University in 1998, master degree in the Computer Engineering Department at Istanbul in 2010, and PhD degree in the Computer Engineering Department at Yildiz Technical University in 2015. He has been a lecturer in the İzmir Kâtip Çelebi University during the years of 2015-2016. Now he is an academician in Istanbul Rumeli University. His major areas of interests are: Image Processing, Machine Learning, Meta Learning and Ensemble Methods.

Shaira OSMANI was born in Kabul Afghanistan in 1994. She received her bachelor's degree from computer engineering department of İzmir Kâtip Çelebi University, in 2016. She has won the gold medal with the project related with this research article in the competition of the Young Brain New Ideas (Genç Beyinler Yeni Fikirler, GBYF). The 5th GBYF Graduation Projects has been organized in the Dokuz Eylül University Tinaztepe campus with the participation of total 11 Aegean Region Universities. Now she is a computer engineer and data analyst in Afghanistan.