Journal of Education for Pure Science- University of Thi-Qar
Vol.11, No1 (June, 2021)
Website: jceps.utq.edu.iq *Email: jceps@eps.utq.edu.iq*

DOI: http://doi.org/10.32792/utq.jceps.11.01.07

Video Summarization for Surveillance System Using key-frame Extraction based on Cluster

Amal H. Awadh[1] **Hazeem B. Taher[2]**

[1,2]Dept. Of Computer Science College of Education for Pure Sciences

Received 11/11/2020 Accepted 10/02/2021 Published 10/06/2021

Abstract:

The amount of data has grown in recent years due to the use of a vast number of videos, which requires time to access them in addition to the difficulty of browsing and retrieving the video content. To fix this issue, it was proposed that the videos be summarized for easy access and that the content of the videos is browsed easier. The primary objective of the video summary is to provide a simple description of the video by removing the redundancy and extracting keyframes from the video. This paper will clarify the four ways that are using to summing up the video based on the keyframe extraction. In frames extraction, the first two methods rely on the threshold value, while the second two methods rely on clustering to extract the keyframes.

Keywords: video summarization, keyframe extraction, absolute difference, Delaunay triangulation, histogram, thresholding.

1. INTRODUCTION:

Digital video technology is increasingly growing. Because of advancements in technology, recording a huge volume of videos becomes very easy. There is a huge bulk of digital content available, such as news, movies, sports, and documentaries. Besides, the need for surveillance has increased significantly due to an increase in security demand [1].

Thousands of surveillance cameras can be seen at public facilities, public transit, hospitals, airports, etc. resulting in vast quantities of information that is impossible to analyze in real-time. Even, it's not that easy to store massive volumes of video files. Rapid processing and effective browsing of the immense amount of data are very important because end users want to access all important facets of data. In this respect, video summarizing plays an important role, as it lets the user search and recover through a wide sequence of videos [1].

Video summarization approaches aim to extract the key events, situations, or objects in a clip and provide a synopsis that is easy to understand. Video summarization's main purpose is to provide a straightforward video interpretation by eliminating duplications and extracting keyframes from the video. Video Summarization is a method of generating and providing a concise descriptive interpretation of the video as a whole within a limited period [2]. There are various methods for video summarization in[4], but it is divided essentially into two distinct forms of Video Summarization strategies. Which is a summary static of video (also known as keyframe extraction) and the other is a dynamic summary(also known as video skim) of the video. Static video storyboard summary includes a compilation of keyframes from the original video and there is no constraint with the issue of time and sequence whereas dynamic video extracts the most appropriate, small, dynamic portions of audio and video to create a summary of video [2,3].

٢.RELATED WORK:

Tommy Chheng in 2007. [10] suggests an automatic algorithm for recognizing a video's specific segments. Using the k-means clustering, the video segments are divided, and he uses Euclidean distance with RGB histogram of the corresponding segment as the distance metric to construct a video description. Particularly targeted at low-quality media, specifically YouTube videos.

Sandra Eliza et al in 2011. [6] suggesting an approach to video summarization. In the first step, pre-sampling is done using one frame per second on video frames. Then color feature in the HSV space is extracted from the video frames. Clustering is then applied to frames and picked from each cluster's keyframe. Finally, another step occurs in which the keyframes are compared using a color histogram to remove those similar keyframes in the summaries generated.

Karim M. Mahmoud et al in 2013. [7] the proposed new method is used Clustering dependent on density. Video frames are pre-sampled with one frame per second rate, in the first stage. In the second stage, video frames extract color features and texture features. The extraction of the color feature is performed in HSV color space using the color histogram. The extraction of the Texture feature is achieved using Discrete Wavelet Transformation (DWT) [8]. Then DBSCAN requests the clustering for format. Lastly, select keyframes from every cluster.

Mr. Satvik Khara et al in 2015. [9] proposed a new method for video summarization using Clustering. Video frames are pre-sampled with one frame per second rate, in the first stage. In the second stage, video frames extract color features, texture features, and Shape features. The extraction of the color feature is performed in HSV color space using the color histogram. The extraction of the Texture feature is achieved using Gabor techniques. Extraction of the Shape feature using Fourier techniques. Then DBSCAN requests the clustering for format. Lastly, select keyframes from every cluster.

Smt. M. Tirupathamma in 2017. [13] the video summarization approach was proposed so that the mainframes were extracted using frame difference. This approach is computationally simple and calculates the number of keyframes dynamically, and has a high precision rate and low error rate. The threshold value is dynamic and will shift during each recording so that keyframes can be efficiently retrieved. The extracted keyframes can represent the video content satisfactorily.

3.THE PROPOSED TECHNIQUES:

Video Summarization is a description that provides an approximate interpretation of the original video series, which can be used for video browsing and retrieval. Various approaches are used in choosing keyframes. These approaches are based mostly on features such as color histogram, histogram, or cluster-related features such as Delaunay clustering and direct method.

3.1 Histogram-based Video Summarization:

This method was conducted in two stages in which the first use frame rate to sample the video after that measure the sample frame threshold (Td)**:**

$$\text{Td} = \mu_{had} + \sigma_{had} * 1.5 \qquad \textbf{(1)}$$

Where μ_{had} is the mean of the histogram of the absolute difference between consecutive image frames, and σ_{had} is the standard deviation of the histogram of the absolute difference between consecutive image frames. In the second stage, the extract the keyframe from the sample image where the absolute differences between consecutive image frames are determined if the difference is greater than the threshold, the current frame is chosen as the keyframe, return this step until processing all sample frames.

Algorithm

Input: Digital Video A.

Output: Summary video.

```
    Step1: Start.
    Step2: Read a video A.
    Step3: Extract video frames.
    Step4: Open the object file for the video writer to write Summarized data from the video.
    Step5: Sample video frames sequence using a frame rate.
    Step6:  Convert to gray.
    Step7: Calculate the absolute difference of the histogram of consecutive image frames.
    Step8: Calculates the mean and standard deviation of absolute difference.
    Step9: Compute Threshold(Td).
    Step10: If the absolute difference is greater than the threshold($d_{hf(i,i+1)} >$ Td) between the adjacent frames, consider the frame as a keyframe and add it to the object file of the video writer, Figure (1).
    Step11: Else go to Step5.
    Step12: Close the writer object file when all sample frames of the input video are read and the difference is computed.
    Step13: The extracted keyframes are combined to form a new video that is the summary video of the given input video.
    Step14: End.
```

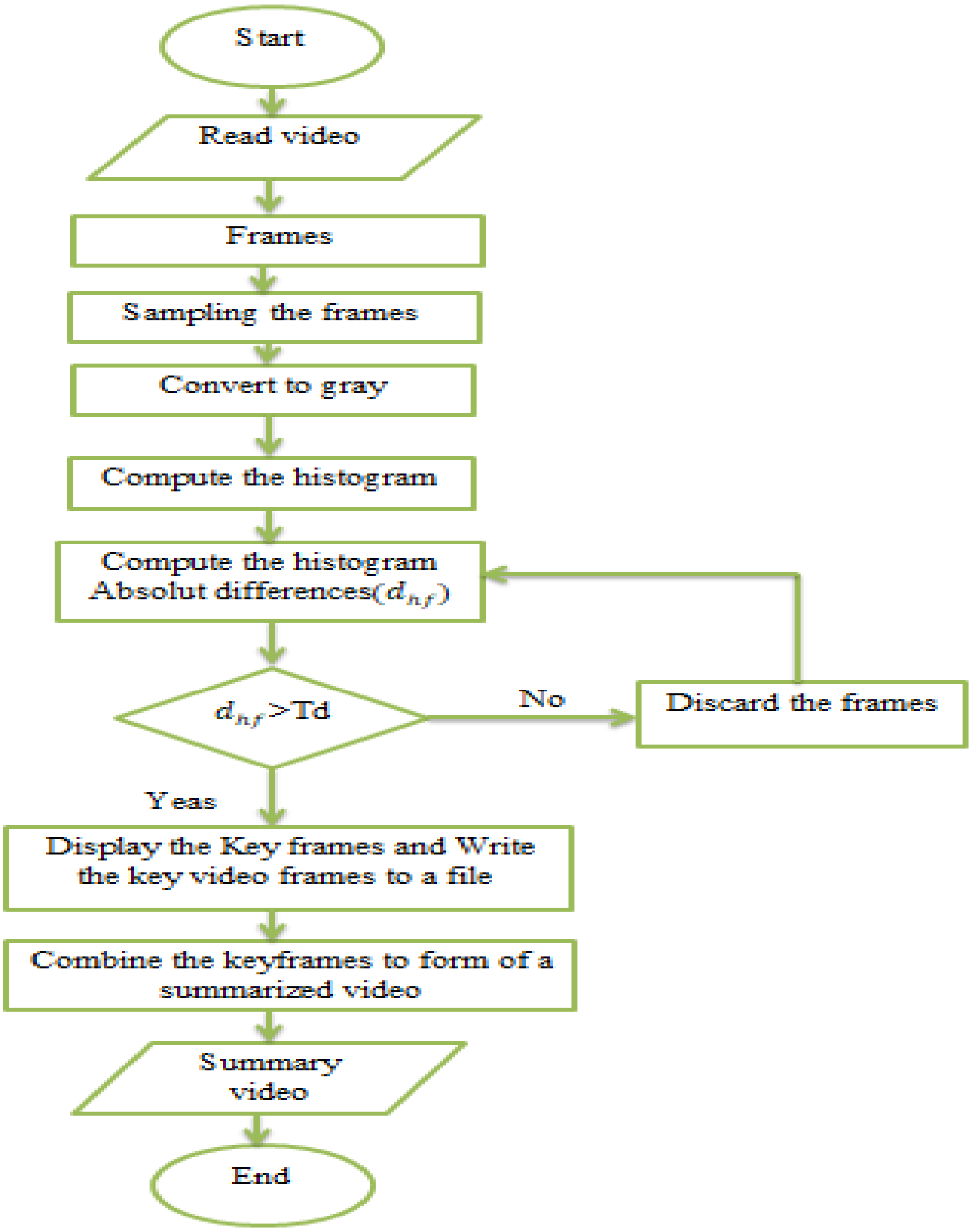

Start
Read video
Frames
Sampling the frames
Convert to gray
Compute the histogram
Compute the histogram Absolut differences(d_{hf})
d_{hf}>Td
No
Discard the frames
Yeas
Display the Key frames and Write the key video frames to a file
Combine the keyframes to form of a summarized video
Summary video
End

Figure (1) Histogram-based Video Summarization

3.2 Color Histogram-based Video Summarization

This process is often carried out in two-stages, In the first stage uses a color histogram for each sample frame, which is using mean and standard deviation to measure threshold(Td):

$$\text{Td} = \mu_{chd} + \sigma_{chd} * 4 \qquad \textbf{(2)}$$

Where μ_{chd} is the mean of the color histogram of the absolute difference between consecutive image frames, and σ_{chd} is the standard deviation of the color histogram of the absolute difference between consecutive image frames. The second stage is to obtain keyframes by comparing each frame with the threshold done by using a color histogram for each frame to estimate the mean which is the color histogram of the absolute difference between the consecutive frame from sampling frames if the means greater than the threshold, that's mean this frame is a keyframe.

Algorithm

Input: Digital Video A.

Output: summary video.

Step1: Start.

Step2: Read a video A.

Step3: Extract video frames.

Step4: Open the object file for the video writer to write Summarized data from the video.

Step5: Sample video frames sequence, the sampling rate is one in every 5 frames that gives us a 6 fps sample for a 30 fps video.

Step6: The first frame in the original video directly is considered the first frame in the summary video.

Step7: Compute the color histogram for the sample frames.

Step8: Calculate the absolute difference of the color histogram of consecutive image frames.

Step9: Calculates the mean and standard deviation of absolute difference.

Step10: Compute Threshold(Td).

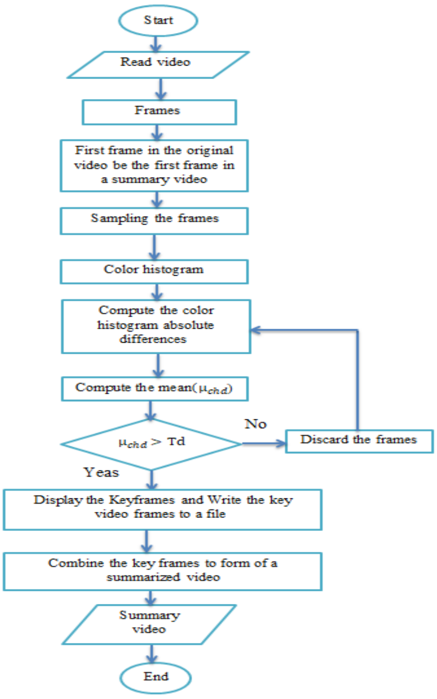
Step11: When the mean of the color histogram absolute difference of consecutive image frames is higher than the threshold(loc_{mean}> Td), accept the frame as a keyframe and add it to the video writer's object file.

Step12: Else go to Step5.

Step13: Close the writer object file when all sample frames of the input video are read and the difference is computed.

Step14: The extracted keyframes are combined to form a new video that is the summary video of the given input video, Figure (2).

Step15**:** End


Start
Read video
Frames
First frame in the original video be the first frame in a summary video
Sampling the frames
Color histogram
Compute the color histogram absolute differences
Compute the mean(μ_{chd})
μ_{chd} > Td
No
Discard the frames
Yeas
Display the Keyframes and Write the key video frames to a file
Combine the key frames to form of a summarized video
Summary video
End

.

3.3 Delaunay Clustering-based Video Summarization

The suggested approach consists of five major steps : (a) pre-sampling of video frames; (b) extraction of features; (c) Principal Component Analysis (d)Delaunay clustering; (e) keyframe extraction, Figure (3).

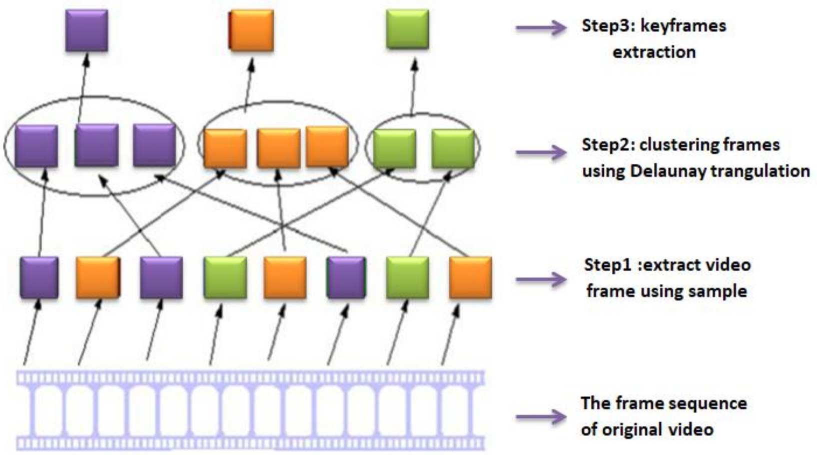

Step3: keyframes extraction
Step2: clustering frames using Delaunay trangulation
Step1 :extract video frame using sample
The frame sequence of original video

Figure (3) Delaunay Clustering-based Video Summarization

Algorithm

Input: Digital Video A.

Output: summary video.

a) **pre-sampling of the video frame**

Sample the video input series at a fixed rate of 5 frames per second.

b) **extraction of features**

each frame represent in HSV space then take Histogram Equalization for them, each frame represents 256-dimensional vector features and the video sequence is described by the matrix $\bar{N}$ for this vector features.

c) **Principal Component Analysis**

the Principal Component Analysis (PCA) is using to reduce matrix dimensions but still obtain the meaning of the frame and to helps in reducing processing time[12].

Steps for PCA algorithm

- find the size of the input matrix $\bar{N}$
- comp1ute the mean value (ΨI) of the matrix.

$$\Psi\mathrm{I}=\frac{1}{n}\sum_{i=1}^{n} a_i \qquad (3)$$

- normalize by subtracting mean.

$$\mathrm{I}\Phi_i = a_i - \Psi\mathrm{I} \qquad (4)$$

Will have the matrix $\bar{M}$={ $\mathrm{I}\Phi_1, \mathrm{I}\Phi_2, \mathrm{I}\Phi_3, \dots, \mathrm{I}\Phi_N$}

- calculate the covariance matrix.

$$C_{ov}=\frac{1}{n}\sum_{i=1}^{n} \mathrm{I}\Phi_i . \mathrm{I}\Phi_i{}^{T} \quad , \; C_{ov} = \bar{M}.\bar{M}^{t} \qquad (5)$$

- calculate eigenvalues (D) of the covariance matrix.

$$\det(\ \bar{M} - DI) = 0 \qquad (6)$$

where I is the Unit or Identity Matrix.

- calculate eigenvectors (V) of the covariance matrix.

$$(\bar{M} - DI)V = 0 \qquad (7)$$

Most $\bar{M}V = \bar{M}D$.

- Sort eigenvalues in descending order by first diagonal sing eigenvalue matrix.
- Put eigenvectors to correspond with eigenvalues.
- In the end, data projection is normalized onto the Eigenspace.

---

d) Delaunay clustering

A triangulation of a set of points P in the plane is a triangulation of the Delaunay if there is no point in P inside the circumcircle of any triangle in the triangulation. The Delaunay triangulation is unique for every group of points P in the plane if there are no four points in P being co-circular (e.g. the vertices of a rectangle). There is no Delaunay triangulation for a set of points that lie in a straight line[11].

Steps for Delaunay clustering

- finding Euclidean distances d(ax, ay) between images vector.

$$d(ax,ay)=\sqrt{\sum_{i=1}^{n}(ax_i - ay_i)^2} \qquad (8)$$

- The mean edge length for each point P_i is known as Local_ML(P_i).
    - ○ Local_ML(P_i)= $\frac{1}{d(P_i)}\sum_{j=1}^{d(P_i)} |e_j|$ (9)
    - ○ Where $|e_j|$ refers to the length of Delaunay edges incident to P_i and $d(P_i)$ refers to the number of Delaunay edges incident to P_i.
- Local SD(P_i) denotes the local standard deviation of the edge length to P_i.

$$\text{Local SD}(P_i)= \sqrt{\frac{1}{d(P_i)}\sum_{j=1}^{d(P_i)}(\text{Local\_ML}(P_i) - |\, e_j|)^2} \qquad (10)$$

- Global_SD(P) denotes the mean of the local standard deviation of all edges.
    - ○ Global_SD(P)= $\frac{1}{N}\sum_{i=1}^{N}$ Local SD(P_i) (11)
    - ○ Where N is the total points number, and where P is the points set.
- Sh_Edge(P_i) refers to a short edge (intra-cluster edge).

Sh_Edge(P_i) = { e_j || e_j|< Local_ML(P_i)− Global_SD(P)} (12)

- S_Edge(P_i) refers to a separating edge (inter-cluster edge).

S_Edge(P_i)= { e_j || e_j |> Local_ML(P_i)+ Global_SD(P)} (13)

- Find and delete the separating edge in the Delaunay diagram for each cluster.

---

a) Keyframe extraction

After deleting all the separating edges in each cluster will extract the keyframe where a keyframe is a frame that is nearest to the center of each cluster. A video summary is a collection of these keyframes.

3.4 Direct Method Video Summarization

Here does not rely on pre-stored videos for summary purposes in the direct summary method, but rather it relies on surveillance cameras when photographing incidents and situations. Instead of photographing and preserving the incident and then summing it up as with the previous techniques that used with this process when the recording is over, will get two videos, one of which is the original and the other is the summary, Figure (4). The algorithm used here is identical to the previous algorithm, as they use Delaunay triangulation in the process of tire assembly. The video presented here will include, as in the original video, the time and date it was registered on.

Algorithm

Input: Set of frames taking from a surveillance camera M.

Output: Recorded video ,summary video.

```
    Step1: Start.
    Step2: Read frames M.
    Step3: Open two object files for the video writer to write a Summarized video and original video
    Step4: Taking one frame out of every five frames.
    Step5: Convert them to gray.
    Step6: Find a histogram for each gray frame.
    Step7: Apply Principle Component Analysis(PCA) to reduce the dimensions.
    Step8: Apply Delaunay triangulation to cluster the frame.
    Step9: Extract the keyframes and companies them to write it to the object file which represents the summary video and the other object file content all the sequence frame which represent the original video.
    Step10: Extract the original and the summary videos at the same time and date of register the frame from the camera.
    Step11: End.
```

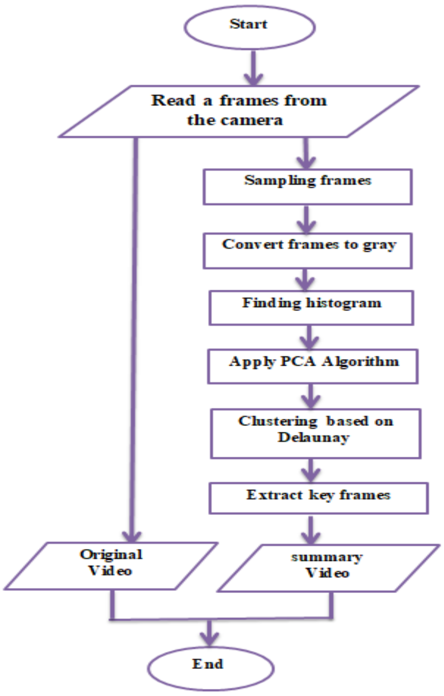


Figure (4) Direct Method Video Summarization

The first approach uses a frame rate for sampling frames unlike other third approaches, where the uses a fixed rate to sample frames. The histogram is used in the first and fourth method, and in the second method, the color histogram is used, while in the third method, it uses the Histogram Equalization. For frame extraction, the first and second methods focus on a certain threshold value, while the third and fourth methods rely on frame extraction clusters. In the first method, the absolute difference is compared with the threshold value, while the mean difference is compared with the threshold value in the second method. Unlike other approaches that rely on captured and stored videos, the fourth approach depends on videos that are recorded directly from the camera.

4.Experimental Results:

To illustrate the summary process, Table (1) includes, a group of videos downloaded from YouTube with the length of each video and the number of frames it contains in addition to the results of each of the first three methods and the number of keyframes extracted from each method.

Table (2) includes the videos taken by the camera, along with the length of each video and the number of frames it includes, as well as the summary results of the fourth method and the number of keyframes extracted by that method.

Original videos			The result of summary videos using					
			1-Histogram-based		2-Color-based		3-Delaunay-clustring	
Video sequence	Total No. of frames	Video Length (h:m: s)	No. of key-frames	video Length (h:m: s)	No. of key-frames	video Length (h:m: s)	No. of key-frames	Video Length (h:m: s)
Video1	752	00:00:30	11	00:00:01	118	00:00:03	30	00:00:01
Video2	878	00:00:35	13	00:00:01	26	00:00:01	67	00:00:02
Video3	2643	00:01:45	37	00:00:01	244	00:00:08	118	00:00:04
Video4	2650	00:01:46	42	00:00:02	200	00:00:06	141	00:00:04
Video5	3548	00:01:58	78	00:00:02	96	00:00:03	62	00:00:02
Video6	5801	00:03:13	134	00:00:04	458	00:00:15	204	00:00:07
Video7	7618	00:04:00	246	00:00:08	60	00:00:02	215	00:00:07
Video8	7461	00:04:09	110	00:00:03	194	00:00:06	242	00:00:08
Video9	7652	00:05:06	233	00:00:07	561	00:00:18	273	00:00:09
Video10	12536	00:08:21	197	00:00:06	15	00:00:01	420	00:00:14
Video11	16234	00:09:01	523	00:00:17	1145	00:00:38	457	00:00:15
Video12	16747	00:09:18	547	00:00:18	1199	00:00:39	471	00:00:15
Video13	18487	00:10:17	608	00:00:20	1284	00:00:42	475	00:00:16
Video14	20021	00:11:08	551	00:00:18	33	00:00:01	685	00:00:22
Video15	21776	00:14:31	770	00:00:25	195	00:00:06	882	00:00:29
Video16	26133	00:14:32	835	00:00:27	1774	00:00:59	930	00:00:31
Video17	29715	00:20:39	1280	00:00:42	2034	00:01:07	958	00:00:32
Video18	48624	00:27:00	1519	00:00:50	465	00:00:15	1428	00:00:47
Video19	49547	00:27:33	1412	00:00:47	3358	00:01:51	1676	00:00:55
Video20	55198	00:30:41	1796	00:00:59	3742	00:02:04	1330	00:00:44

Original Video			Summary Video	
Video sequence	**Video length (hh:mm:ss)**	**Total No. of frame**	**Video length (hh:mm:ss)**	**No. of keyframes**
Video1	00:00:19	596	00:00:01	23
Video2	00:00:25	752	00:00:01	40
Video3	00: 00:30	909	00:00:03	99
video4	00: 00:33	1010	00:00:03	99
video5	00: 00:37	1130	00:00:03	91
Video6	00:00:47	1410	00:00:02	54
Video7	00:01:09	2097	00:00:03	79
Video8	00:01:12	2166	00:00:03	84
Video9	00:01:21	2436	00:00:03	79
Video10	00:01:24	2535	00:00:03	102
Video11	00:01:47	3222	00:00:05	137
Video12	00:01:47	3231	00:00:03	119
Video13	00:03:15	5870	00:00:07	201
Video14	00:03:47	٦٩٥٦	00:00:07	٢٢٧
Video15	00:03:57	7123	00:00:08	259
Video16	00:04:59	8999	00:00:07	207
Video17	٠٠:٠٦:٥٦	١٢٤٩٦	00:00:13	٤١٧
Video18	٠٠:٠٨:٥٥	١٦٠٥٧	00:00:15	٤٥٧
Video19	٠٠:١٠:٠٨	١٨٢٦٦	00:00:24	٧٣٣
Video20	٠٠:١٠:٤٦	١٩٤٠٢	00:00:24	٧٣٦
Video21	٠٠:١١:٤٤	٢١١٣٢	00:00:26	٧٨٥
Video22	00:12:36	٢٢٦٩٤	00:00:22	٦٧٣
Video23	00:13:58	25154	00:00:25	776
Video24	00:15:07	27236	00:00:31	959
Video25	00:31:38	٥٦٩٤٠	00:00:52	١٥٨٢
Video26	00:36:23	65508	00:00:55	1677

5.CONCLUSIONS AND FUTURE WORKS:

Have been suggested four methods of summarizing the videos in this paper, based on the keyframe extraction. The first and second methods are techniques that focus on the threshold value in collecting keyframes to summarize images, while the third and fourth methods that focus on clustering are two techniques that are free of user-specified modeling criteria and produce video summaries by collecting in fewer frames the visual quality of the original videos than other summarization techniques.

The findings of the first approach revealed that it is the quickest in the process of frame processing and keyframe extraction also shows that the keyframe extraction dependent on the color histogram is that, in terms of visual quality, this technique is the best relative to the other techniques. The third approach resolved the problems of the previous threshold value-based methods due Some videos will not display any frames or very few frames that might exist when the threshold value is very high, resulting in a very low summary quality of this video. The direct method helped to get rid of much of the difficulties found with the previous methods, including issues relating to bad storage and storage precision, as the output varies with the same video stored in two different resolutions, so bad storage does not impact the direct method because it provides direct results and the imaging quality is the same as the quality of the summary. It also helped to decrease the time required in the previous methods to process the videos.

In future work to extract a key-frame, use the texture feature with a color histogram. also may analyze other physical characteristics, such as edge and motion descriptors, and their fusion, and use them to extract keyframes, to get more meaningful video summaries, and also use a more comprehensive set of characteristics such as color, motion, texture, and shape, together with a successful feature fusion strategy. The Delaunay clustering can be extended to multiple features such as text, audio, and motion features to cluster video frames.

References:

[1] Muhammad Ajmal, Muhammad Husnain Ashraf, Muhammad Shakir, Yasir Abbas, and Faiz Ali Shah "Video Summarization: Techniques and Classification", International Conference on Computer Vision and Graphics, pp. 1–13, (2012).

[2] Dharmesh Tank "A Survey on sports video summarization”, International Journal for Science and Advance Research in Technology - Volume 2 Issue 10, October 2016.

[3] Himani Parekh, and Pratik Nayak "A Survey on KeyFrame Based Video Summarization Techniques”, International Journal of Engineering Research in Computer Science and Engineering (IJERCSE), Vol 4, Issue 11, (November 2017).

[4] Z. El, Y. Tabii, and A. Benkaddour "Video Summarization: Techniques and Applications”, International Journal of Computer and Information Engineering, vol. 9, no. 4, pp. 882–887, (2015).

[5] T. Sebastian and Jiby J. Puthiyidam, "A Survey on Video Summarization Techniques", International Journal of Computer Applications, vol. 132, no. 13, pp. 31–33, (December 2015).

[6] De Avila, S. E. F., Lopes, A. P. B., da Luz Jr, A., & de Albuquerque Araújo, A. (2011). "VSUMM: A mechanism designed to produce static video summaries and a novel evaluation method". Pattern Recognition Letters, 32(1), 56-68.

[7] Mahmoud, K. M., Ismail, M. A., & Ghanem, N. M. (2013). "Vscan: an enhanced video summarization using density-based spatial clustering". In International conference on image analysis and processing (pp. 733-742). Springer, Berlin, Heidelberg.

[8] Stanković, R. S., & Falkowski, B. J. (2003), "The Haar wavelet transforms its status and achievements", Computers & Electrical Engineering, 29(1), 25-44.

[9] Khara, S., Modi, B., Shah, D. J., & Thakkar, R. (2015). "Video Summarization using clustering". International Journal of Innovative Research in Technology, 2(6), 31-36.

[10] Chheng, T. (2007). "Video summarization using clustering". Department of the Computer Science University of California, Irvine.

[11] Manuel D. Salas, "A Shock-Fitting Primer", (2009), ISBN: 978-1-4398-0758-3.

[12] I.T. Jolliffe, "Principal Component Analysis", Second Edition, 2002, ISBN: 0-386-95442-2.

[13] Tirupathamma, S. M. (2017). "Key frame-based video summarization using frame difference". International Journal of Innovative Computer Science & Engineering, 4(3).