Compare Between Histogram Similarity and Histogram Differencing For More Brief Key Frames Extraction from Video Stream

Ekhlas Falih Naser  
E-mail: 110022@uotechnology.edu.iq  
Computer Sciences Department, University of Technology/Baghdad/Iraq

Abstract. Speedy increase in the employ of the content of multimedia is very commonly perceived in existing generations. Video is one of the extreme exceedingly recovered data of multimedia which can provide quick fixes and immediate gratification for whole the kinds of user information willingness. Detection of Key frame from Videos and classification of videos are the necessary protocols in the complements tasks of video retrieval. The aim of this research is to compare between histogram similarity and histogram differencing for more brief key frames extraction from video stream. The suggested system for key frame extraction has three steps. Firstly, the video frames series will take and the characteristics for points of interest are elicited utilizing SUSAN detector. After eliciting interest characteristics from all video frames images, secondly, K-Means clustering technique was utilized for these features to construct the clusters with a number of interest points. Thirdly, a histogram builds for each video frame based on numbers of features in each cluster. X-axis of a histogram represents the cluster number and y-axis represents the number of features in each cluster. For key frames extraction, numbers for each cluster can be offered and a query histogram can be constructed based on entered clusters' numbers. A query histogram was matched with every video frame histogram using Manhattan distance to discover the similar histogram to query histogram. After chosen similar histogram, it can extract all key frames from that video frames. The experimental results show that the number of key frames extracted from the video is very brief when using the histogram similarity compared to the histogram differencing.  

1. Introduction
Classification of video is the procedure of gathering videos depend on contents' similarities. Classification of video became important in the existing technology trends to contend the diversity of user requirements [1]. The challenging agents of video classification involve number of videos, size of the videos, formats and various categories of videos. The classification of video guide the user towards extra wished for selecting by joining the same videos' kind together. Video content is one of the necessary discriminative operators for classification [2]. Videos are essentially constructed from frames and it is clear that frames extracted directly from videos are comprised of lots of repeat frames as same frame is casted from 1to 2 or extra seconds of duration time. It is extremely necessary to recognize the salient frames which can convey the key characteristics of the videos which is achieved via detection of key frame [3]. The removal of repeated frames from videos is usually describes as detection of key frame. The detection of key frames is extra challenging due to videos' length and this requires the effective algorithms for key frames identification. Key frames are candidate frames realizing the salient videos contents uniquely without any repetition or overlap [4]. Detection of videos' key frames minimizes the calculation burdens which required in videos' processing and enables also classification in optimal way. Key frame detection consists of various of advantages for real time such as compression of video, abstraction of video, searching and browsing, summarization of text from videos, generation of story, generation of automatic index form and different other achievements in the field of retrieval techniques[5].

2. Related Works

There are various methods for key frames extraction as illustrated bellow:

2.1 Unsupervised Clustering _ Shot Boundary Detection
Zhuang had suggested a technique for extraction key frame from videos by employing unsupervised clustering technique [6]. The clustering is undergoing towards segmented of the shots in a video to gather the key frames in a singular shot in one class or more classes. The clustering is achieved depend on the visual frames' content via histogram of a frame as a threshold within the shot. The method's efficiency utilized is validated with several videos in real world. However, the technique is calculated heavy due to inclusion differencing of histogram as crucial parameters for gathering of key frames into clusters.

2.2 Statistical Techniques Based Key Frame Extraction
Retrieval of videos rely on frames' objects has increased for the time being due to quick simply accessibility to several resources online. Sivic introduce a technique utilizing view point regional descriptors of the images' objects for extraction the key frame [7]. Moreover, the frames ranking is implemented rely on the spatial shape of the frames' objects. The experimentation is executed on the videos of entertainment, especially two movies.

2.3. Histogram Differencing
Enabling arrival quickly to the visual videos' content is one of the major objectives for the extraction technique of key frames. Gianluigi had achieved extraction of the key frame from the frames by employing the technique of histogram differencing. Key frames can be extracted via discovering the curvature points in the curve of accumulated frame differences. The histogram differencing needs massive calculations for frames' processing to determine the key frames [8].

2.4 Robust key frame extraction method
Yang [9] has offered a quick and strong technique is relying on low level of frames' features, such as features of structural and features of color. A 2-steps technique is utilized to discover accurate key frames which suit the whole video sequence's content. For the first step, a replace sequence is acquired rely on the difference in characteristics of color among adjacent frames from the original sequence. For the second stage, via analyzing the feature of structural differences among adjacent frames in the replace sequence, the eventual key frame sequence is acquired. Then, optimization is done rely on overall number of eventual key frames to confirmed the success extraction of key frame.

3. Proposed Methodology

The suggested methodology for key frame extraction has three steps. Firstly, the video frames series take and the characteristics of interest points are elicited by employing SUSAN detector. After eliciting interest characteristics from all video frames images, secondly, K-Means clustering technique was utilized for these features to construct the clusters with a number of interest points. Thirdly, a histogram builds for each video frame based on numbers of features in each cluster. X-axis of a histogram represents the cluster number and y-axis represents the number of features in each cluster. For key frames extraction, numbers for each cluster can be offered and a query histogram can be constructed based on entered clusters' numbers. A query histogram was matched with every video frame histogram using Manhattan distance to discover the similar histogram to query histogram. After chosen similar histogram, it can extract all key frames from that video frames. Figure 1 displays block diagram of the system for key frame extraction.
3.1 SUSAN Corner Detection algorithm

SUSAN (“Smallest Unvalue Segment Assimilating Nucleus”) approach was offered by circular-mask containing a center-pixel which called (nucleus). If the brightness of a pixel-mask will compare with the nucleus' brightness, then the mask's zone will introduce which contain a similar (identical) to the nucleus' brightness. This mask's zone will know as "USAN" is derived from words “Univalue Segment Assimilating Nucleus” [10]. The function among each mask's pixel and mask's nucleus was utilized to detect a corner. Eq. 1 utilized for comparison.

\[
\text{comp}(\text{pix}, \text{pix}_0) = \begin{cases} 1, & \text{if } |\text{img(pix)} - \text{img(pix}_0)| < \text{thr}; \\ 0, & \text{otherwise} \end{cases}
\]

Where pix0 is the coordinates of nucleus's pixel; pix is the pixel coordinates of other points within a mask; \(\text{comp}(\text{pix}, \text{pix}_0)\) appears the outcome of a comparison task; \(\text{Img(pix)}\) appears pixel point for gray rate; thr appears the threshold for gray difference's that utilized for determining the characteristic domain ability dimension and minimum contrast which will discover via SUSAN approach [10]. The comparison can happen for each pixel within a mask. To calculate the total \((n)\) , Eq. 2 was utilized.

\[
\text{n(pix}_0) = \sum_{\text{pix}} \text{comp(pix, pix}_0) 
\]

The total \((n)\) offered the number of pixels in (USAN) and will give the zone of USAN. \((n)\) is compared after that to a specific threshold \((g)\) which is set to be exactly a half of \(n_{\text{max}}\) due to the actuality that USAN zone must be lower than half of the zone's mask if the nucleus sites within corner and must be local minimum. To construct an initial corner-response, Eq.3 will employ.

Figure 1: Block diagram of the proposed system for key frames extraction
$$Rc(pix_0) = \begin{cases} 1, & g - n(pix_0) \text{ if } n(pix_0) < g; \\ 0, & \text{otherwise} \end{cases}$$  
(3)

Rc (pix_0) presents an initial corner-response [10].

### 3.2 Features Clustering using K_Means Algorithm

The clustering of K_Means represents a partition-based-cluster-analysis technique. This clustering kind was considered one of the extreme methods for unsupervised classification. Also, it is one of the extreme unsupervised learning techniques due to its simplicity. A classifier of unsupervised case that a training data-set with recognized labels of the class is required for the former to train the classification principle, while this training data-set and the class labels knowledge in the data-set are not required finally[11]. In this technique, an iterative style was utilized for data-base clustering. It gather the number of required clusters and an initial average as (inputs) and outcomes in latter as (output). In the situation that the technique is required to outcome (K )clusters then there would be (K )initial mediums and (K )final mediums . K-means results (K )final mediums in completion that grant the answer to why the technique was called K_Means. The first stage of K_Means depends on chosen of (k )objects as initial cluster-centroids, then calculates a distance between each centroid within a cluster and each one of the objects ,allocate it to the nearby one of the clusters, change the averages of every cluster and then reiterate this style. The optimal resolution for frequent repetitions will take as a final solution. An easier simple technique is the problem initialization via a random chosen of (k) data-points from the specific data. The reminder of the data-points undergo classification to (k) clusters within the distance [11]. The centers are afterwards modified through a computation of the centers in the (k) clusters.

### 3.3 Histogram Similarity

The speed of feature's matching was accomplished via a unique stage of indexing relied on the Manhattan's amount of clustering amounts. Calculate the distance would be traveled via a function of Manhattan to gain with one data-point to other if the grid was followed. Manhattan distance for two components is the summation of differences of their corresponding elements [12]. A formula for a point \(X=(X_1, X_2, \text{etc.})\) with a point \(Y=(Y_1, Y_2, \text{etc.})\) can be calculated via Eq.4.

$$d = \sum_{i=0}^{n} |x_i - y_i|$$  
(4)

### 3.4 Suggested Algorithm

The suggested algorithm was offered as bellow

| Input : video stream, Histogram based on numbers of features in each cluster |
| Output : Key frames for brief video representation |

begin

**Step1:**  
1) Enter AVI video and Covert video stream into frames (imgs)
2) Enter the test numbers for each cluster (test)

**Step 2:**  
For i=1 to No. of Frames
2.1 Detect the interest points for (imgs[i]) using SUSAN corner
2.2 Apply the algorithm of k-means clustering for all interest points of imgs[i] to obtain clusters which contain the number of interest points in each cluster
2.3 Construct a histogram based on the number of interest points in each cluster
End For

**Step 3:** //For Key frames Extraction
Enter the number of features for each cluster and find a histogram of these clusters.
Display all Frames that matching the histogram using Eq.4.

**Step 4:** Show all key frames as brief video representation

End.
4. Empirical Results

The empirical results of proposed methodology were discussed and displayed in this section. A proposed methodology was implemented using C# programming language. Ten kinds of data bases are utilized to evaluate the proposed methodology. The images within a data base are colored with size of 320 × 240 pixels. The proposed methodology consisting of multiple steps:

1) At the first step, loading the video stream and the frames are extracted as shown in the figure (2) for car video and figure (7) for scoter video.

![Figure 2: Car video, a) input video b) extraction frames from car](image)

2) The interest points are detected in the second step using SUSAN corner detector as exhibited at figure (3) for car dataset and figure (8) for scoter dataset.

![Figure 3: Interest points detection from car dataset using SUSAN corner detector](image)

3) After eliciting interest features from all video frames images, the clustering approach that utilizing K_Means was employed for these features to construct the clusters with a number of interest points as exhibited at figure (4) for car dataset and figure (9) for scoter dataset.
Figure 4: K-Means clustering on car dataset

4) A histogram builds for each video frame based on numbers of features in each cluster as illustrated in figure(5). X-axis of a histogram represents the cluster number and y-axis represents the number of features in each cluster. For key frames extraction, numbers for each cluster can be offered and a query histogram can be constructed based on entered clusters’ numbers as exhibited at figure (5) for car dataset and figure (10) for scoter dataset.

Figure 5: Histogram builds for a sample key frame of car dataset

5) A query histogram was matched with every video frame histogram using Manhattan distance to discover the similar histogram to query histogram. After chosen similar histogram, it can extract all key frames from that video frames with similar histograms as exhibited at figure (6) for car dataset and figure (11) for scoter dataset.

Figure 6: Samples of key frames which are extracted based on similar histogram of car dataset
Figure 7: Scoter video, a) input video b) extraction frames from Scoter

Figure 8: Interest points detection from scoter dataset using SUSAN corner detector

Figure 9: K-Means clustering on scoter dataset
Figure 10: Histogram builds for a sample key frame of scoter dataset

Figure 11: Samples of key frames which are extracted based on similar histogram of scoter dataset

Table (1) shows the comparison result between histogram similarity and histogram differencing for samples of video in term of more brief key frames extraction from video stream

Table 1: comparison result between histogram similarity and histogram differencing

| Video Name | Total Numbers of Frames | Number of Key frames which are extracted based on histogram similarity | Number of Key frames which are extracted based on histogram differencing |
|------------|-------------------------|---------------------------------------------------------------------|-------------------------------------------------------------------|
| Car        | 250                     | 70                                                                  | 180                                                              |
| Scoter     | 333                     | 92                                                                  | 241                                                              |
| Tracery    | 120                     | 30                                                                  | 90                                                               |
| Airhorse   | 50                      | 7                                                                   | 43                                                               |

5. Conclusion

The aim of this research is to compare between histogram similarity and histogram differencing for more brief key frames extraction from video stream. A suggested methodology utilized the detector of SUSAN to elicited interest points. SUSAN algorithm has the advantages of being robust to noises and rotation of the image because its mask is circular and used statistic method to count univalue points. Also, a suggested methodology utilized K-means clustering technique to group points of interest into clusters. K-means is fast and uncomplicated technique for clustering data. As illustrated in Table1, the experimental results show that the number of key frames extracted from the video is very brief when
using the histogram similarity compared to the histogram differencing. For example, the number of key frames extracted from a car video is 70 when using the histogram similarity, while it is 180 when using the histogram differencing; also the number of key frames extracted from a scooter video is 92 when using the histogram similarity, while it is 241 when using the histogram differencing.

References

[1] Karpathy A, Toderici G, Shetty S, Leung T, Sukthankar R and Fei-Fei L 2014 Large-scale video classification with convolutional neural networks. In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, 1725-1732.

[2] Rani N, Neethu O. and Ponnath N 2015 Automatic Vehicle Tracking System Based on Fixed Thresholding and Histogram Based Edge Processing. International Journal of Electrical and Computer Engineering, 5(4).

[3] Pawar M, Perianayagam L and Rani, N 2017 Region based image classification using watershed transform techniques. International Conference on Intelligent Computing and Control (I2C2), 1-5.

[4] Rani N and VasudevT 2015 Automatic detection of telugu single and multi-character text blocks in handwritten words. International Conference on Computing and Network Communications (CoCoNet), 234-240.

[5] Lei S, Xie G and Yan G 2014 A novel key-frame extraction approach for both video summary and video index. The Scientific World Journal. Volume 2014, Article ID 695168.

[6] Pramod N, Paul C and Shobha N 2018 Key Frame Extraction from Videos- A Comparative Study. Journal of Advanced Research in Dynamical & Control Systems, Vol.10, No.12.

[7] Shah K and Makwana R 2017 Shot Boundary Detection using Logarithmic Intensity Histogram: An Application for Video Retrieval. International Journal of Applied Engineering Research, 12 (22), 12616-12624.

[8] Shi Y, Yang H, Gong M, Liu X and Xia Y 2017 A fast and robust key frame extraction method for video copyright protection. Journal of Electrical and Computer Engineering, Volume 2017, Article ID 1231794, 7 pages.

[9] Zong Z, and Gong Q 2017 Key frame extraction based on dynamic color histogram and fast wavelet histogram. IEEE International Conference on Information and Automation (ICIA),183-188.

[10] Pattar S 2015 Study of Corner Detection Algorithms and Evaluation Methods International Journal of Innovative Research in Science, Engineering and Technology, Vol. 4, No. 5.

[11]Ahhammad F 2015 A modified K-Means Algorithm for Big Data Clustering. B.Sc. Dissertation. Computer Science and Information System; IBAIS University; Dhaka, Bangladesh.

[12] Karim A and Nasser E 2018 Image Retrieval from Video Streams Databases using Similarity of Clustering Histogram. Al-Mansour Journal, No.29, pp.1-22.