REVIEW ON THE VIDEO SEGMENTATION TECHNIQUES
Piyush Bisht

Dept. of Computer Science & Engineering, Amity School of Engineering and Technology,
Amity University Uttar Pradesh, Noida, India

ABSTRACT:

From the fast few years video summarization has become the active area of research. With the increasing load of the multimedia content on the internet, it has become significantly important. To extract desired content from video various techniques have been introduced for the video segmentation as it is the integral part of the video summarization like color histogram differences[25][27], edge change ratio[27], standard deviation of pixel intensities[27] etc. One video segmentation technique is not good for all the types of videos. Some techniques give better result in one case and some in the other case. Another issue encountered in segmenting the video into shots are their space and time complexity.

Keywords: Color Histogram, Frames, Shot Segmentation, Thresholding, Video Summarization

[1] INTRODUCTION:

Video summarization has become the latest area of research in recent years due to increase in demand for the multimedia content on the internet. It helps the user to find the video on the internet in the short time stem. It plays important role in video indexing and video retrieval. In this the shot detection is the first step. There are number of techniques being introduced for the shot detection such as color histogram differences[25][27], edge change ratio[27], standard deviation of pixel intensities[27]. The shots are the elementary units of the video which are
further processed to get the key frames. They are called the basic units of video which are the continuous number of the frames captured from a camera for a particular series of time[5]. Shots may be classified in the following category:

- Hard cuts
- Fades
- Dissolves

Video detection algorithms[25] always have the difficulty of finding the shot boundaries because of the object motion, variations in the illumination[25]. Visual coherence technique[14] is the very less explored in the last few years. While doing the video segmentation, the main challenge is to find the edits correctly. Video summarization is mainly divided into two categories that is static summarization[1] and video skimming[1]. In static video summarization[1], we generally combined the still images to form the summarised video and in dynamic skimming[1] we combine the moving clips to form the summarised video. Static video appears to be less interesting as they have still images and occupy less space which makes them flexible. Key frames are extracted from the video frames using their features like RGB channel, color histograms etc[1]. Among all the approaches for shot segmentation, shot clustering[2] is the appropriate approach as it has the low computational cost. The shots is composed of hundreds of images which may have different content. The color histogram[25] method is one of the best to select from for the video segmentation as method. One of the major problem that existed while segmenting the videos into shots is the threshold, we have to carefully select the threshold value as it may be responsible for the true hit and false hit while segmenting process[14]. Precision and recall may be affected by the threshold value. Choosing one threshold value is infeasible so we use dynamic threshold[19].

[2] LITERATURE REVIEW:

[2.1] COLOR HISTOGRAM DIFFERENCES:

The color histogram-based shot boundary[25][27] detection algorithm is one of the most reliable variants of histogram-based detection algorithms. Its basic idea is that the color content does not change rapidly within but across shots. Thus, hard cuts and other short-lasting transitions can be detected as single peaks in the time series of the differences between color histograms of contiguous frames or of frames a certain distance k apart[27].

The color histogram-based shot boundary[25][27] detection algorithm is a standout amongst the most solid variations of histogram-based detection algorithms[25][27]. Its fundamental thought is that the color content does not change quickly inside but rather crosswise over shots. Subsequently, hard cuts and other transition can be distinguished as single crests in the time arrangement of the contrasts between color histograms of adjoining frames or of edges a specific separation k separated.

[2.2] EDGE CHANGE RATIO:

Let sn be the number of edge pixels in frame n, and the number of entering and exiting edge pixels in frames n and n-1, respectively.
ECRn = \max \left( \frac{E(x,y,t)}{E(x,y,t-1)} \right)

Hard cuts are perceived as segregated tops; during fade ins/outs the quantity of approaching/active edges prevails; and during a dissolves, at first the active edges of the principal shot just before the approaching[27]. Edges of the second shot begin to overwhelm the second 50% of a dissolve.

[2.3] STANDARD DEVIATION OF PIXEL INTENSITIES:

During video creation fades are delivered by a monotone and generally direct scaling of the pixel intensities after some time[27]. This intensity scaling is obviously unmistakable in the time arrangement of the standard deviation of pixel intensities as portrayed.

\[ s(E(x,y,t)) = \sqrt{\mu(X^2) - \mu^2(X)} \]

The intensity scaling is straightforwardly shown in the standard deviation of the pixel intensities. The scaling function utilized amid video generation and the standard deviation of the pixel intensities are indistinguishable with the exception of a steady element.

[2.4] EDGE-BASED CONTRAST:

Autonomous of the sort of scaling function an onlooker watches lost differentiation and sharpness of the pictures during a dissolve that for the most part achieves its greatest centre of the dissolve[27]. Henceforth, the essential thought of the along these lines characterized edge-based contrast is to catch and stress the contrast and/or sharpness to empower dissolve detection.

\[ EC(K) = 1 + (s(K) - w(K) - 1) / (s(K) + w(K) + 1), \quad EC(K) \in [0, 2] \]

[2.5] FADE AND DISSOLVE BOUNDARY DETECTION:

The technique is not ready to recognize smooth shot boundaries as fades and dissolves. As said beforehand, these kind of shot changes show a triangular-like shape in the convolved shading histogram distinction signal HistDifconv[i], because of the straight move between shots that video altering impacts as fades and dissolves present[25]. the strategy is not ready to distinguish smooth shot limits as fades and dissolves.

Choosing whether a cut exists between two back to back frames is typically done by looking at components of the two frames utilizing an frame, while different works depend on clustering procedures. Fixed thresholds for the most part can't bargain well with various sorts of videos[19]. While clustering performs ineffectively[19] when the distinctions computed around likely cuts are not high contrasted with the mean of the distinctions of non-cut frames, since gathering the casings in two clusters (typical frames and cut frames) could prompt miss an expansive number of shots. The errand of relegate labels to substance is normally left to clients who transfer clips, making issues of vagueness, absence of data, and interpretation.
The video is at first fragmented into shots by the recognizable proof of camera cuts. Every shot is then spoken to by one or more frames and its length and movement substance is evaluated. In the initial steps of the algorithm, a color similitude measure of shots is registered, which is called Backward Shot Coherence (BSC)[15]. Valleys are found in BSC measures and are identified some Potential Scene Boundaries (PSC)[15]

The histogram contrast between two continuous edges surpasses some limit, a shot boundary is recognized. At the point when the color pictures are viewed as, some proper weights are allocated to the histogram of every color part contingent upon the significance of color space, so weighted histogram based correlation was proposed. In proposed a shot detection technique by breaking down edge change proportion (ECR)[18] between back to buck casings. The rate of edge pixels that enter and exit between two back to back edges are ascertained. The cuts and progressive moves can be identified by contrasting the proportion of entering and leaving edge pixels. The impediment of the edge based strategies is that they neglect to deliver great results when the video succession contains fast question movements. To conquer this deficiency, an algorithm for movement pay was produced. The square coordinating methodology[18] is utilized and movement estimation is performed. Nonetheless, movement based methodology is computationally expensive. Once the sudden moves are distinguished, then it is critical to identify the continuous shot moves, for example, fade in and fade outs which are available in the video for outlining a compelling video retrieval system.

It is watched that the connected component based techniques are easy to actualize yet are not hearty since they accept that the regular elements, for example, shading or dark power are shared by the content pixels having a place with the same associated area. Despite what might be expected, composition based strategies[3] might be unacceptable for little textual styles and poor difference content. Also, the rate of execution is quick for edge and angle based strategies however they create all the more false positives when the foundation is complex.

Movement based sampling[4] of frames has likewise been utilized for video summarization. The fundamental downside of these techniques is that they catch neighbourhood varieties and might miss critical sections while longer portions may seem various times with comparative substance. Clustering is a natural solution, where the whole information is isolated into a pre-characterized number of cluster (in light of the synopsis length) trailed by the choice of models from every cluster. The clustering plans[4] show great execution by and large; be that as it may, they might wind up selecting summary frames just from the overwhelming clusters and might neglect fascinating occasions which happen occasionally.

[3] METHODOLOGY:

Steps of scene segmentation[14]:
a)Key-frames selection; b) shot coherence calculation and potential scenes creation; c) Similar adjacent scenes removal; d) Motion similarity comparison and; d) High-similar scenes removal[14].
In the above figure, steps involve in video summarization are given: a) Input video b) Frame extraction c) Histogram comparison of frames d) Shot detection on the basis of the threshold e) Reassmbling of the frames in the backward direction f) Formation of the summarised video from the frames.

a) Input video: In this step different type of video is selected to be summarised. Video taken should not have resolution more than 360p and its length should not be more than five minutes.

b) Frame Extraction: In this step we extract the frame from the video taken as the input. Frame extraction process depends on the frame of the video.

c) Histogram comparison of the frames: Histogram of the frames are formed and compared with histogram of the next frame.

d) Shot detection on the basis of the threshold: After the histogram are compared then after that on the basis of the threshold the shot is detected. If the color difference value is more than the threshold value then we consider that point as the cut.

e) Reassembling of the frames in the backward direction: On the scattered frames algorithm is applied to select the key frames for the video summarization.

f) Formation of the summarised video from the frames: Summarised video is formed with the frames selected in the last step.
REVIEW ON THE VIDEO SEGMENTATION TECHNIQUES

It computes the similitude utilizing the histogram's convergence, considering two casings comparable when they introduce no less than 95% of histogram crossing point[14]. After the main key-frame be chosen, more key-frames are chosen hunting down the following casing which is most comparable with every other frame of the shot. After the key-frames were picked, our system computes the shot cognizance of a shot with different past shots. This intelligibility quality is called as Backward Shot Coherence Weighted (BSCW)[14]. The BSC is a shot cognizance esteem got looking at the key-outlines in a window of shots. The BSCW, nonetheless, is BSC esteem which was duplicated by a quality called Temporal Memory” [14].

Each separated components contrast qualities are contrasted and the predefined threshold at various stages. As the initial step, sum of absolute differences (SAD)[20] is computed between the two sequential frames and is contrasted and the threshold T1. On the off chance that the distinction is under T1 the comparing edge is not a cut frame and the remaining preparing is skipped. Conversely, if SAD is over the limit, contrast in light of square coordinating is ascertained [20]. Feature extraction is performed utilizing the HSV predominant color parts of the frames taking into account the piece wise histogram contrasts that are figured for continuous casing. The separated component firmly fulfills the two properties of the video hard cut recognition. The separation between any two sequential casings having a place with the same shot is much littler than the separation between any two continuous frames fitting in with various shots. Video frames have close histogram features while a different frame compares to very separated element values. The Bin wise histogram distinction between squares of the two continuous frames is ascertained utilizing the similitude estimations.

The procedure produces static outlines of recordings by correlation between visual features to concentrate key frames. Visual points of interest and many-sided quality of a video can’t be measured by one single element itself, so more elements are expected to totally comprehend the video with the goal that key casings can be separated all the more productively. SSIM or Structural Similarity Index[15] is a technique for measuring comparability between two frames. It gauges picture quality taking into account introductory uncompressed or distortion free picture as reference. Connection contrast measures whether two casings are equivalent in light of color substance. Relationship distinction measures whether two edges are equivalent taking into account shading content. On the off chance that just color histogram is utilized then even the casings having comparable color content yet distinctive visual substance will be dealt with as comparable. Snippet of Inertia aides fit as a fiddle portrayal. SSIM has esteem 1 for the indistinguishable pair of edges. The SSIM is intended to enhance conventional measurements like PSNR and MSE, which have ended up being conflicting with human eye recognition. A picture minute is the weighted normal of picture pixel's intensities [15]. Color histogram just measures the color circulation. On the off chance that there are two unique pictures yet with same color circulation then histogram measure will distinguish them as comparable pictures however this is not the situation with color minutes. Normalized histograms are produced for every frame in the shot, and after that a whole of histograms is computed position by position. The outcome is partitioned by the quantity of casings in the shot. A pixel is considered fringe on the off chance that it is on the picture outskirt or if nothing else one of its four neighbours (up, down, left and right) has an alternate color. A pixel is viewed as inside if its four
neighbours have the same color. After pixels order, a color histogram is processed considering just fringe pixels and another is figured considering inside pixels. A threshold for determination of shot change frame is a vital component in the method of SCD[23]. When all is said in done, if the elements of a video frame are bigger than a threshold, then we view the casing as a shot change. In this manner, we decide a video outline as a shot change frame, however not as per the set up edge. The proposed technique uses a normal of ADH inside of augmentation sliding window[23].

A video division strategy utilizing a histogram-based fuzzy c-means (HBFCM) clustering algorithm[26]. We propose a video segmentation method utilizing a histogram-based fuzzy c-means (HBFCM) clustering algorithm. The shot change (SC) group, the suspected shot change (SSC)[26] bunch, and the no shot change (NSC) group. Key frames are chosen from every shot. The HBFCM clustering algorithm has the accompanying two particular benefits: to start with, it needn't bother with the threshold as required by the shot change recognition approach; second, it presents the SSC group which is not considered in the bunching approach. The HBFCM clustering algorithm has the accompanying two unmistakable benefits: in the first place, it needn't bother with the edge as required by shot change discovery calculations; second, it presents the SSC cluster which is not considered in other clustering algorithms.[26]

Luminance Moments[17] are characteristic picture properties and two sorts are viewed as valuable here. The calculating so as to shade normal luminance can be effectively gotten the mean of R, G, and B channels of the considerable number of pixels. One point of preference of this element is that the R, G and B parts are specifically computed as opposed to embracing measurement decrease as done in piece color histogram estimation. Both the normal luminance and difference luminance are very valuable and they are consolidated together as Luminance Moments Feature[17].

The histograms are registered for all the three channels R, G and B independently. The aggregate histograms are gotten from every histograms. We fit a straight line to the combined histogram. At that point the line parameters, for example, incline and y-intercept are computed. Thus, it results in six component highlight vector for every video frame. This element vector is thusly used to distinguish the shot moves in view of the frame contrasts. To distinguish fade moves, we consider just the casing contrasts comparing to non-cut frames, i.e. the fade detection is trailed by cut-detection process.[18]

A key frame is extricated taking into account spatio-temporal color distribution. The color appropriations of the Y ,I and Q parts of a casing is spoken to by its shading minutes and likelihood conveyance is particularly portrayed by its minutes. The transiently most extreme event frame is built by considering the spatial and worldly circulation of the pixels all through the video shot. The movement in recordings regularly cause haziness and low difference in the video frame. During this procedure, we utilize the Contrast Limit Adaptive Histogram Equalization(CLAHE)[3] technique to improve the picture. the Adaptive Histogram Equalization(AHE)[3] enhances the difference in pictures however contrasts from normal histogram evening out. The AHE tends to over enhance clamor in generally homogeneous
districts of a picture. Histogram Equalization(HE) functions admirably when the appropriation of pixel qualities is comparative all through the picture. Notwithstanding, when the picture contains locales that are essentially lighter or darker than the greater part of the image.

Video streams have an innate repetition among them because of the high frame rate of camcorders. Along these lines, a great summary strategy ought to concentrate on recognizing the basic occasions in a given video. Notwithstanding, if two summary frames independently outfit imperative, however copy data, then the net data picked up by selecting them two is not maximal. A metric figuring the repetition among an arrangement of chose frames is in this way of fundamental significance. A rundown system taking into account amplifying the thorough and totally unrelated criteria guarantees that the embodiment of the first video is caught well in the summary (comprehensive) and that there is negligible duplication of data.

The distinguished timestamps for the scene limits, the recordings will be delayed consequently toward the end of every scene and the video specialists quickly will be provoked to choose one of the three applicant key-frames of every scene (initially, center and last casing of every scene) as the potential key-frame delegate utilizing the gave graphical client interface. We endeavour to quantify the pertinence level of every video scene to a gathering of pre-characterized abnormal state visual ideas. Furthermore, the end-clients' needs and inclinations into any of those ideas can be removed specifically. Joining these two arrangements of information can help with deciding a specific end-client level of enthusiasm into a particular video scene. Every video scene is connected with a key-frame that has the most astounding potential to speak to that specific video portion outwardly and semantically. The competitor outline that has the most elevated choice rate (among the three conceivable applicants) for every scene will be chosen as the delegate key-frame for that specific video section. Filter highlights for the greater part of the chose preparing pictures nearby the key-frame are extricated for visual similarity testing. The acquired visual elements from the key-frame are differentiated two by two against those components recovered from the chose sub-class to recognize any potential matches between these two arrangements of information.

[4] SUMMARY:

According to review done the color histogram difference algorithm is one of the best algorithm that is being introduced. It depends on the idea that the color changes rapidly during the the change of the shots.Color histogram difference can easily detect hard cuts but can not detect dissolves and fades. Edge change ratio algorithm depends on the idea of change of the sharpness and contrast. Its computational time is more as comparison to the color histogram difference but more effective than it. It can not be used to detect the dissolves and fades.Standard deviation of the pixel intensity is another method for the shot detection. It assumes that in the single shot intensity does not change from on frame to another. Edge based contrast algorithm measure the loss of the contrast and sharpness in the frames.This algorithm can easily detect the dissolve which very difficult to detect.
REFERENCES


REVIEW ON THE VIDEO SEGMENTATION TECHNIQUES


