Keyframe Selection of Frame Similarity to Generate Scene Segmentation Based on Point Operation

**Wisnu Widiarto 1, 2, Mochamad Hariadi 1, Eko Mulyanto Yuniarno 1**

**1**Department of Electrical Engineering, Sepuluh Nopember Institute of Technology, Surabaya, Indonesia

**2**Informatics Department, Sebelas Maret University, Surakarta, Indonesia

|  |  |  |
| --- | --- | --- |
| **Article Info** |  | **ABSTRACT** |
| ***Article history:***Received Jul 21th, 2017Revised Dec 1th, 2017Accepted Feb 18th, 2018 |  | Video segmentation has been done by grouping similar frames according to the threshold. Two-frame similarity calculations have been performed based on several operations on the frame: point operation, spatial operation, geometric operation and arithmatic operation. In this research, similarity calculations have been applied using point operation: frame difference, gamma correction and peak signal to noise ratio. Three-point operation has been performed in accordance with the intensity and pixel frame values. Frame differences have been operated based on the pixel value level. Gamma correction has analyzed pixel values ​​and lighting values. The peak signal to noise ratio (PSNR) has been related to the difference value (noise) between the original frame and the next frame. If the distance difference between the two frames was smaller then the two frames were more similar. If two frames had a higher gamma correction factor, then the correction factor would have an increasingly similar effect on the two frames. If the value of PSNR was greater then the comparison of two frames would be more similar. The combination of the three point operation methods would be able to determine several similar frames incorporated in the same segment. |
| ***Keyword:***Frame-differenceGamma-correctionPeak-signal-to-noise-ratioPoint-operationSimilarity |
| *Copyright © 2018 Institute of Advanced Engineering and Science. All rights reserved.* |
| ***Corresponding Author:***Wisnu Widiarto, Department of Electrical Engineering, Sepuluh Nopember Institute of Technology (ITS) SurabayaKampus ITS Sukolilo, Surabaya, Indonesia 60111Email: wisnu13@mhs.ee.its.ac.idInformatics Department, Sebelas Maret University (UNS) SurakartaJl. Ir. Sutami No.36A Surakarta, Indonesia 57126Email: wisnu.widiarto@staff.uns.ac.id |

1. **INTRODUCTION**

At this time, the management of data and documents is very important, including the management of video document. The video document is divided into two main layers: shot and scene [1], [2]. Video is a collection of frames arranged in sequence. The frames are sequences of events arranged by shot and scene. The video document is usually managed and analyzed based on four basic points which are the basic structure of the video hierarchy: frame, shot, scene, and video sequence [3], [4].

Video management and video analysis creates several management techniques that are categorized into two types: static (keyframe) and dynamic (skimming) [5], [6]. Static management techniques are done by selecting prominent frames and important to be used as key frames [7]. Then, several key frames are selected from each segment to be collected and rearranged. The key frame (static) represents certain frames, the keyframe collection is a collection of selected frames and prominent from the video scene. Dynamic management techniques are done by selecting some of the video sequences that are considered important, so the video becomes shorter. Skimming (dynamic) is an abstraction of moving frames and contains video segments of the video scene.

In this research, a similar frame generation will be done. The method used is a method based on the operation of the point, the usual operation applied to the frame. The point operation method used is: frame difference, gamma correction and peak signal to noise ratio. The frame similarity calculation is done on the pixel value between the initial frame and the next frame, for all pixel values. The calculated value is based on RGB color values (Red, Green, and Blue).

The similarity measures are applied to many applications, such as video management to analyze and detect shots or scenes. Some of researcher proposed a method for understanding scene using three steps: segmentation, object detection and motion features or combining classification, annotation and segmentation [8]. One of the purposes of video analysis is to detect a shot or scene. Scene is a collection of shots that have relationships and create a storyboard from the video [9]. Scene can be detected using several processes: using crossentropy for two histograms [10], using the maximum entropy method (MEM) based on linear prediction [11], using the hidden markov model [12], using markov chain monte carlo [13], similarity based on color and motion information using shot similarity graph (SSG) [1], using the normalized graph cut approach (NCut) [14], edge detection based on mathematical morphology [15], using second generation curvelet transforms [16].

Shot is the basic element that forms a video. Shot is a sequential frame, recorded from a camera [2], [11], [17]. Shot can be detected using transitions between successive frames in a video. Shot detection techniques have been done based on the color histogram [18], based on pixel differences [19], and motion information [20].

Key frames are sub-sections of a video that can represent the content of the video and can provide information about the video using fewer frames. The main purpose of the key frame is to make the video shorter than the original video without reducing the core information, by minimizing the number of frames and eliminating the frame redundancy [21]. The selection of key frames can be done with three methods: cluster based (grouping similar frames into one group, then taking multiple frames from each group), energy minimalization based (minimized using looping techniques), and sequential based (creating new key frames in the scene different) [22]. The Cluster-based frame selection method has been done by selecting frames to represent each cluster [23], [24]. The selected frame is called a key frame.

1. **RESEARCH METHOD**

The key frame is the selected frame of a video that can represent important video content. Users can watch video content by displaying highlights from key frame. Key frame extraction techniques can be classified in three ways: sequential comparison of color, color clustering, and sum of frame-to-frame differences [25]. In this research, key frame generation will be applied using frame similarity process based on point operation. Point operation is closely related to various image processing techniques, including frame differences, gamma correction and peak signal to noise ratio.

A frame consists of several pixels that have color information values in numerical form. These numerical values can be presented in 8bit x 3 integers. Each frame has a different color value. The difference of the frame value will cause the difference of distance and the difference of pixel value between two frames. The general form of distance difference between two frames using Euclidean distance is (1).

 (1)

C = color (R,G,B), i = pixel, MxN = size of frame,

PF = pixel value of the selected frame, PF’ = pixel value of the next frame

The numerical values of frame pixels can be related to actual lighting. That relationship is called gamma. The difference value between two frames (selected frames and actual lighting) is called gamma correction. In this research, the selected frame is considered to be a frame that has not been affected by gamma correction (F) and the next frame is assumed as a frame that has been affected by gamma correction (F'). In general, the form of gamma transformation can be written in (2).

 (2)

m = frame number, i = pixel, C = color (R,G,B), F = the frame selection (without gamma correction),

F’ = the next frame (with gamma correction), ∂ = gamma correction (0 < ∂ <1).

There are two criteria for determining the similarity assessment of two frames: objective fidelity criteria and subjective fidelity criteria. Objective fidelity criteria can be done by creating a mathematical function to calculate the difference and similarity of the two frames. For the case of frames with MxN size, the mean square error value (between the initial frame and the next frame) will be represented by (3).

 (3)

M = the length of the frame (row), N = the width of the frame (column),

= the pixel value of the initial frame, = the pixel value of the next frame

If the next frame is compared with the initial frame then there is an error or noise signal, peak signal to noise ratio denoted by PSNR using (4). If the MSE value is lower then the two frames are more similar. If the PSNR value gets bigger then the two frames are getting more similar.

 (4)

m = frame number, i = pixel, C = color (R,G,B),

MSE = the mean square error value, PSNR = the peak signal to noise ratio

The purpose of this research is to generate key frame of a video using frame similar process. This research is done in three stages as shown in figure 1: streaming, processing, and generating. The streaming stage is the stage to separate streaming between streaming video and streaming audio. Then, the video streaming results are divided into frames and the audio stream results are not used in the operation process (removed). The second stage is the processing stage which is divided into four levels: operation, segmentation, selection, and deletion. Processing stage is the stage to measure frame similar, classify frame similar, select key frame and remove frame redundancy. The generating stage is the last step to collect the key frames that have been obtained from the selection of key frames in the selection process for the processing stage, as follow in figure 1.

A comparison between two successive frames will determine the similarity value of the two frames. The similarity scores for all frames will be calculated and used to determine the position of the segments of similar frames. The combination of three methods (frame difference, gamma correction, and psnr) will give the conclusion that two frames have similar states, dissimilar, hesitations and dissolve, as shown in table 1. The frames located on each segment are classified into two types: candidates and redundancy. The candidate is the selected frame as the key frame candidate. Redundancy is a frame that is classed as a frame similar and selected to be removed.



Figure 1. Frame work of research method

Table 1. The Combination result of point operation

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Frame** | **FD** | **GC** | **PSNR** | **Combine** | **Conclusion** | **Key frame** | **Key frame Selection** |
| 1-2 | S | S | S | S | SegmenF1-F5 | CandidateKey Frame | **Key frame****Selection** |
| 2-3 | S | S | S | S |
| 3-4 | S | S | S | S |
| 4-5 | S | S | S | S |
| 5-6 | D | D | D | D | SegmenF6-F9 | CandidateKey Frame | **Key frame****Selection** |
| 6-7 | S | S | S | S |
| 7-8 | S | S | S | S |
| 8-9 | S | S | S | S |
| 9-10 | S | S | D | SS | SimilarHesitation | SegmenF10-F12 | **Key frame****Selection** |
| 10-11 | S | D | S | SS |
| 11-12 | D | S | S | SS |
| 12-13 | S | D | D | DD | DissimilarHesitation | **Delete** | **Delete** |
| 13-14 | D | S | D | DD |
| 14-15 | D | D | S | DD |
| 15-16 | D | D | D | D | Dissolve | **Remove** | **Remove** |
| 16-17 | D | D | D | D |
| 17-18 | D | D | D | D |
| FD = frame difference, GC = gamma correction, PSNR = peak signal to noise ratioS = similar, D = dissimilar, SS = similar hesitation, DD = dissimilar hesitation |

1. **RESULTS AND DISCUSSION**

In this case, the experimental video is a video that has a size of 102,715 KB and the time length of 00:08:38. Video is divided into 6220 frames and each frame has a size of 2292 x 1667 pixels. All frames use three RGB color parameters (Red, Green, and Blue). Each frame (Fm) is compared with the previous frame (Fm-1) and the next frame (Fm+1), except the initial frame (F1) and the last frame (Fn). The frame comparison corresponds to the color parameter (RGB). The comparison between two frames uses three-method operations: frame difference, gamma correction, and psnr.

The combination result of three methods is used to determine the value of similarity. If the calculation results give a similar value in sequence, then the new segment will be identified until the value of dissimilar is obtained. In the segment, the frame is selected as the key frame candidate. The key frame candidate will be the key frame. The separator between a segment and the next segment is the one-time dissimilar state.

If three combinations of point operation produce two similar values and one dissimilar value then it is called a similar hesitation. If the combined result is a similar hesitation then the frame remains in the segment, and the frame is considered as a redundancy frame and then deleted the frame. If a combination produces only one similar value (two methods produce a dissimilar value) it is called dissimilar hesitation. The combination of three methods that generated dissimilar hesitation will provide a recommendation that the frame is worth to be removed.

If there is a dissimilar state on all methods (the conclusion of three methods: dissimilar) and the condition is experienced in sequence at least three times, called dissolve. Dissolve will cause the event that the frame was removed from the candidate key frame. The selected key frame candidates will be used as key frames (each segment). All key frames will be collected and used to represent the frame of a video.

All frames are calculated as similarity values. Each similarity calculation uses the same three method operation. The comparison result between the two frames (examples: frame #0370 – frame #0377) is shown in table 2 (frame difference), table 3 (gamma correction) and table 4 (pnsr). The three-method application creates different combination values for different frame comparisons, examples as follow in table 1. If the three comparative methods produce similar conditions then the combination is called similar. If the three-method operation cause two frames in a dissimilar condition then it is said dissimilar.

The comparison is done, the initial frame (example: **frame #0373**) and the next frame (**frame #0374**), starting from the first point (cell (1, 1)) until the last point (cell (M, N)). Frame difference value of frame #0373 and frame #0374: Red = 1.88655; Green = 2.00988; Blue = 2.00575; Average = 1.96739; Result = Similar1. Gamma correction value of frame #0373 and frame #0374: Red = 0.95098922; Green = 0.95996575; Blue = 0.95792674; Average = 0.95629390; Result = Similar2. PSNR value of frame #0373 and frame #0374: Red = 30.59189; Green = 30.24738; Blue = 30.06772; Average = 30.30233; Result = Similar3. Produce of frame #0373 and frame #0374 = Similar1 x Similar2 x Similar3 = **Similar**.

The next comparison is done on all frames for all points (cells) every frame, starting from the first point until the last point. For the second example (dissimilar condition): the initial frame (example: **frame #0374**) and the next frame (**frame #0375**). Frame difference value of frame #0374 and frame #0375: Red = 26.63293; Green = 22.55145; Blue = 18.37646; Average = 22.52028; Result = Dissimilar1. Gamma correction value of frame #0374 and frame #0375: Red = 0.60008055; Green = 0.64589050; Blue = 0.68210545; Average = 0.64269217; Result = Dissimilar2. PSNR value of frame #0374 and frame #0375: Red = 14.25604; Green = 15.28285; Blue = 16.36976; Average = 15.30288; Result = Dissimilar3. Produce of frame #0374 and frame #0375 = Dissimilar1 x Dissimilar2 x Dissimilar3 = **Dissimilar**. The same comparison is done on all frames for the all point operation (cell).

Table 2. Frame similarity value of frame difference (frame #0372 – frame #0377)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **frame** | **fd\_red** | **fd\_green** | **fd\_blue** | **avg\_fd** | **summary** |
| . . . | . . . | . . . | . . . | . . . | . . . |
| #0372-#0373 | 0.84198 | 0.72688 | 0.86451 | 0.81113 | Similar |
| #0373-#0374 | 1.88655 | 2.00988 | 2.00575 | 1.96739 | Similar |
| #0374-#0375 | 26.63293 | 22.55145 | 18.37646 | 22.52028 | Dissimilar |
| #0375-#0376 | 0.54072 | 0.70613 | 0.53263 | 0.59316 | Similar |
| #0376-#0377 | 2.92401 | 2.86417 | 2.96498 | 2.91772 | Similar |
| . . . | . . . | . . . | . . . | . . . | . . . |

Table 3. The value of gamma correction (frame #0372 – frame #0377)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **frame** | **gc\_red** | **gc\_green** | **gc\_blue** | **avg\_gc** | **summary** |
| . . . | . . . | . . . | . . . | . . . | . . . |
| #0372-#0373 | 0.97518991 | 0.98103876 | 0.97658475 | 0.97760447 | Similar |
| #0373-#0374 | 0.95098922 | 0.95996575 | 0.95792674 | 0.95629390 | Similar |
| #0374-#0375 | 0.60008055 | 0.64589050 | 0.68210545 | 0.64269217 | Dissimilar |
| #0375-#0376 | 0.97845369 | 0.98435614 | 0.98028198 | 0.98103060 | Similar |
| #0376-#0377 | 0.92514107 | 0.93969747 | 0.93510159 | 0.93331338 | Similar |
| . . . | . . . | . . . | . . . | . . . | . . . |

Table 4. Psnr value of examples of the frame comparison (frame #0372 – frame #0377)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **frame** | **psnr\_red** | **psnr\_green** | **psnr\_blue** | **avg\_psnr** | **summary** |
| . . . | . . . | . . . | . . . | . . . | . . . |
| #0372-#0373 | 41.25402 | 42.78729 | 41.26261 | 41.76797 | Similar |
| #0373-#0374 | 30.59189 | 30.24738 | 30.06772 | 30.30233 | Similar |
| #0374-#0375 | 14.25604 | 15.28285 | 16.36976 | 15.30288 | Dissimilar |
| #0375-#0376 | 39.53136 | 37.98148 | 39.40038 | 38.97107 | Similar |
| #0376-#0377 | 27.42145 | 27.64012 | 27.96547 | 27.67568 | Similar |
| . . . | . . . | . . . | . . . | . . . | . . . |

In this research, the calculated frame is the first 600 frames taken from the video. The first 600 frames are processed to determine the scene, all the frames located in each scene and the number of frames in each scene. The determination process uses three point operations: frame difference, gamma correction and the peak signal to noise ratio. The frame assignment in each scene raises the number of frames in each scene. Different operating methods can cause different scenes and number of frames in each scene. To overcome these differences, then the rules are determined to obtain the approximate number of scenes and the number of frames per scene. The rules include: if the number of frames in a scene is less than 5 frames, then the scene will be deleted (removed); calculation of min-max difference between three point operation and determined the least difference value; as shown in table 5.

The color histogram, by Widiarto, has been used as a tool for shot detections, several key frames have been selected as representative of each shot on a video [20]. Widiarto has used pixel differences to determine key frame, selected key frames have been used to form comic strips [21]. In this research, the determination of the scene is processed based on the difference of frame between the two closest frames by detecting each frame using a combination of three point operations in order to obtain more accurate segmentation of the scene.

The comparison results of frame similarity calculations for the three methods of operation are shown using table as shown in table 6. The table shows that the difference in mean point (cell) values, obtained from each frame (for each red, green, blue), determines the position of similarity located, in a similar position or dissimilar. If the table shows a sudden change of value, then it is said that the scene change is detected suddenly. If the table shows slow changes, the scene changes will be detected by dissolve. The table is made to show that two frames are in a similar or dissimilar area, for calculations using frame difference (FD), gamma correction calculation (GC), and psnr calculation (PSNR).

Calculation of each frame and comparison of two successive frames has been done then determined the scene and the number of frames in each scene. To define the scene in this research using three parameters (three point operation) then apply the rules in table 5. The final decision of the scene and the number of frames per scene for a combination of the three point operations can be shown as table 6. The first **600 frames** of the video produce **16 scenes** with number of frames respectively (**527 frames**): 132, 17, 16, 33, 33, 32, 32, 12, 34, 30, 28, 21, 25, 30, 31, and 21. **Removed** frames are (**21 frames**): #436 – #441 (6 frames), #495 – #506 (12 frames), and #575 – #577 (3 frames). **Deleted** frames are (**52 frames**): #133 – #136 (4 frames), #154 – #156 (3 frames),#271 – #275 (5 frames), #308 – #310 (3 frames), #323 – #340 (18 frames), #433 – #435 (3 frames), #442 – #448 (7 frames), #507 – #511 (5 frames), #573 – #574 (2 frames), and #578 – #579 (2 frames).

Table 5. Scene number combination of three methods point operation

|  |  |  |  |
| --- | --- | --- | --- |
| **frame number threshold of every scene** | **number of scene****in the operation** | **FD/GC/PSNR** | **distance of min-max** |
| **FD** | **GC** | **PSNR** | **min** | **min** |
| All of frame in the scene is available | 46 | 61 | 124 | 46 | 124 | 78 |
| if the number of frame in the scene < 2 then scene is removed | 23 | 24 | 33 | 23 | 33 | 10 |
| if the number of frame in the scene < 3 then scene is removed | 21 | 21 | 29 | 21 | 29 | 8 |
| if the number of frame in the scene < 4 then scene is removed | 18 | 20 | 23 | 18 | 23 | 5 |
| **if the number of frame in the scene < 5 then scene is removed** | **18** | **19** | **22** | **18** | **22** | **4** |
| if the number of frame in the scene < 6 then scene is removed | 16 | 18 | 22 | 16 | 22 | 6 |

Table 6. Frame number and number of frames for every scene using three methods point operation

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **No** |  | **FD** |  |  | **GC** |  |  | **PSNR** |  | **Combination result (FD/GC/PSNR)** |
| **number of frames** | **frame numbers** | **number of frames** | **frame numbers** | **number of frames** | **frame numbers** | **scene number****(number of frame)** | **frame numbers** | **delete (D)****remove (R)** |
| 1 | 132 | #001–#132 | 132 | #001–#132 | 132 | #001–#132 | **1 (132)** | #001–#132 | - |
| 2 | 21 | #133–#153 | 17 | #137–#153 | 8 | #133–#140 | **2 (17)** | #137–#153 | #133–#136 (D) |
| 3 |  |  |  |  | 13 | #141–#153 |  |  |  |
| 4 | 16 | #157–#172 | 19 | #154–#172 | 16 | #157–#172 | **3 (16)** | #157–#172 | #154–#156 (D) |
| 5 | 33 | #173–#205 | 33 | #173–#205 | 33 | #173–#205 | **4 (33)** | #173–#205 | - |
| 6 | 33 | #206–#238 | 33 | #206–#238 | 33 | #206–#238 | **5 (33)** | #206–#238 | - |
| 7 | 32 | #239–#270 | 32 | #239–#270 | 32 | #239–#270 | **6 (32)** | #239–#270 | - |
| 8 | 37 | #271–#307 | 9 | #276–#284 | 14 | #271–#284 | **7 (32)** | #276–#307 | #271–#275 (D) |
| 9 |  |  | 23 | #285–#307 | 15 | #285–#299 |  |  |  |
| 10 |  |  |  |  | 8 | #300–#307 |  |  |  |
| 11 | 33 | #308–#340 | 26 | #308–#333 | 12 | #311–#322 | **8 (12)** | #311–#322 | #308–#310 (D)#323–#340 (D) |
| 12 | 34 | #341–#374 | 34 | #341–#374 | 34 | #341–#374 | **9 (34)** | #341–#374 | - |
| 13 | 22 | #375–#396 | 30 | #375–#404 | 6 | #375–#380 | **10 (30)** | #375–#404 | - |
| 14 | 8 | #397–#404 |  |  | 12 | #381–#392 |  |  |  |
| 15 |  |  |  |  | 12 | #393–#404 |  |  |  |
| 16 | 31 | #405–#435 | 31 | #405–#435 | 6 | #405–#410 | **11 (28)** | #405–#432 | #433–#435 (D) |
| 17 |  |  |  |  | 22 | #411–#432 |  |  |  |
| 18 |  |  |  |  |  |  |  |  | #436–#441 (R) |
| 19 | 5 | #442–#446 | 24 | #446–#469 | 21 | #449–#469 | **12 (21)** | #449–#469 | #442–#448 (D) |
| 20 | 23 | #447–#469 |  |  |  |  |  |  |  |
| 21 | 5 | #470–#474 | 8 | #470–#477 | 25 | #470–#494 | **13 (25)** | #470–#494 | - |
| 22 | 20 | #475–#494 | 17 | #478–#494 |  |  |  |  |  |
| 23 |  |  |  |  |  |  |  |  | #495–#506 (R) |
| 24 | 5 | #507–#511 | 5 | #507–#511 |  |  |  |  | #507–#511 (D) |
| 25 | 30 | #512–#541 | 30 | #512–#541 | 10 | #512–#521 | **14 (30)** | #512–#541 | - |
| 26 |  |  |  |  | 20 | #522–#541 |  |  |  |
| 27 | 33 | #542–#574 | 33 | #542–#574 | 31 | #542–#572 | **15 (31)** | #542–#572 | #573–#574 (D) |
| 28 |  |  |  |  |  |  |  |  | #575–#577 (R) |
| 29 | 23 | #578–#600 | 23 | #578–#600 | 21 | #580–#600 | **16 (21)** | #580–#600 | #578–#579 (D) |

1. **CONCLUSION**

The generation of key frames with similar processes is based on three point operating methods: frame difference, gamma correction, and peak signal to noise ratio. The combination of the three methods makes the selected key frame more appropriate. The removed frame is more precisely because the redundancy frame selection process uses three parameters. If all three process methods of point operation produce similar conditions, then the two frames are called frame similar and have the same scene position. If three processes show dissimilar conditions, then the two frames is called different pixel values and two frames are in different scene positions. In the case of conditions that make dissimilar occur sequentially, the frames are assumed to be in the disolve area, so the automatic frame removal is done and the frame is not used as a candidate key frame. This research uses a video that is divided into 6220 frames and captured the first 600 frames as research material. The combination of three point operation methods (frame difference, gamma correction and pnsr) produces 16 scenes and each scene has a number of different frames (16 scenes = 527 frames). Frames categorized in the remove/delete classification are redundant and / or dubious frames to serve as a scene, so that the frames are removed/deleted (removed frames = 21 and deleted frames = 52).

**REFERENCES**

1. Z. Rasheed and M. Shah, “Detection and representation of scenes in videos,” *IEEE Transactions on Multimedia*, vol. 11, no. 6, pp. 1097–1105, 2005.
2. Y. Zhu and Z. Ming, “SVM-based video scene classification and segmentation,” *International Conference on Multimedia and Ubiquitous Engineering 2008 (MUE 2008)*, 2008, pp. 407–412.
3. Y. Zhai and M. Shah, “Video scene segmentation using Markov chain Monte Carlo,” *IEEE Transactions on Multimedia*, vol. 8, no. 4, pp. 686–697, 2006.
4. A. Chergui, *et al.*, “Video scene segmentation using the shot transition detection by local characterization of the points of interest,” *2012 6th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT)*, 2012, pp. 404–41.
5. B.T. Truong and S. Venkatesh, “Video abstraction: A systematic review and classification,” *ACM Transactions on Multimedia Computing Communications and Applications (TOMCCAP)*, vol. 3, no. 1, pp. 1–37, 2007.
6. Y. Xue and W. Zhicheng, “Video Segmentation and Summarization Based on Genetic Algorithm,” *IEEE 2011–4th International Congress on Image and Signal Processing*, 2011, pp. 460–464.
7. M.A. Mizher, *et al.*, “A meaningful Compact Key Frames Extraction in Complex Video Shots,” Indonesian Journal of Electrical Engineering and Computer Science, vol. 7, no. 3, pp. 818–829, 2017.
8. L.J. Li, *et al.*, “Towards total scene understanding: classification, annotation and segmentation in an automatic frame work,” *IEEE Conference on computer vision and pattern recognition (CVPR2009)*, 2009, 2036–2043.
9. Z. Xiong, *et al.*, “Semantic retrieval of video - review of research on video retrieval in meetings, movies and broadcast news, and sports,” *IEEE Signal Processing Magazine*, vol. 23, no. 2, pp. 18–27, 2006.
10. S.H. Kim and R.H. Park, “A novel approach to scene change detection using a cross entropy,” *Proceeding 2000 International conference on image processing*, 2000, 937–940.
11. P.Y. Yeoh and S.A.R. Abu-Bakar, “Maximum entropy method (MEM) for accurate motion tracking,” *TENCON 2003 Conference on convergent technologies for Asia-Pasific region*, 2003, 345–349.
12. J. Huang, *et al.*, “Joint scene classification and segmentation based on hidden markov model,” *IEEE Transactions on Multimedia*, vol. 7, no. 3, pp. 538–550, 2005.
13. Y. Song, *et al.*, “MCMC-based scene segmentation method using structure of video,” *International Symposium on Communications and Information Technologies (ISCIT)*, 2010, pp. 862–866.
14. Y. Zhao, *et al.*, “Scene Segmentation and Categorization Using NCuts,” *2007 IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–7.
15. D. Aghlmandi and K. Faez, “Automatic Segmentation of Glottal Space from Video Images Based on Mathematical Morphology and the Hough Transform,” *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 2, no. 2, pp. 223–230, 2012.
16. R.C. Johnson, *et al.*, “Curvelet Transform based Retinal Image Analysis,” *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 3, no. 3, pp. 366–371, 2013.
17. H.B. Kang, “A hierarchical approuch to scene segmentation,” *IEEE Workshop on Content-Based Access of Image and Video Libraries, 2001 (CBAIVL 2001)*, 2001, pp. 65–71.
18. W. Widiarto, *et al.*, “Video summarization using a key frame selection based on shot segmentation,” *2015 International Conference on Science in Information Technology (ICSITech)*, 2015, pp. 207–212.
19. W. Widiarto, *et al.*, “Shot segmentation of video animation to generate comic strip based on key frame selection,” *2015 IEEE International Conference on Control System, Computing and Engineering (ICCSCE)*, 2015, pp. 303–308.
20. M. del Fabro and L. Boszormenyi, “Video Scene Detection Based On Recurring Motion Patterns,” *2010 Second International Conferences on Advances in Multimedia*, 2010, pp. 113–118.
21. G. Guan, *et al.*, “Keypoint-based key frame selection,” *IEEE Transactions on circuits and systems for video technology*, vol. 23, no. 4, pp. 729–734, 2013.
22. C. Panagiotakis, *et al.*, “Equivalent keyframes selection based on Iso-Content Principles,” *IEEE Transactions on circuits and systems for video technology*, vol. 19, no. 3, pp. 447–451, 2009.
23. S.E.F. De Avila, *et al.*, “VSUMM: A mechanism designed to produce static video summaries and a novel evaluation method,” *Pattern Recognition Letters*, vol. 32, no. 1, pp. 56–68, 2011.
24. M. Pournazari, *et al.*, “Video Summarization Based on a Fuzzy Based Incremental Clustering,” *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 4, no. 4, pp. 593–602, 2014
25. T. Liu, *et al.*, “A novel video key-frame-extraction algorithm based on perceived motion energy model,” *IEEE Transactions on circuits and systems for video technology*, vol. 13, no. 10, pp. 1006–1013, 2003.