International Journal of Electrical and Computer Engineering (IJECE)
Vol. 7, No. 5, October 2017, pp. 2565 – 2573
ISSN: 2088-8708, DOI: 10.11591/ijece.v7i5.pp2565-2573

Video Shot Boundary Detection Using The Scale Invariant Feature Transform and RGB Color Channels

Zaynab El khattabi[1], Youness Tabii[2], and Abdelhamid Benkaddour[3]
[1,3]LIROSA Laboratory, Faculty of Sciences, Abdelmalek Essaadi University,Tetuan, Morocco
[2]LIROSA Laboratory, National School of Applied Sciences, Abdelmalek Essaadi University, Tetuan, Morocco

Article Info

Article history:
Received: May 5, 2017
Revised: Jun 12, 2017
Accepted: Jun 29, 2017

Keyword:
Video Segmentation
Shot Boundary Detection
Gradual Transition
Abrupt Change
SIFT

ABSTRACT

Segmentation of the video sequence by detecting shot changes is essential for video analysis, indexing and retrieval. In this context, a shot boundary detection algorithm is proposed in this paper based on the scale invariant feature transform (SIFT). The first step of our method consists on a top down search scheme to detect the locations of transitions by comparing the ratio of matched features extracted via SIFT for every RGB channel of video frames. The overview step provides the locations of boundaries. Secondly, a moving average calculation is performed to determine the type of transition. The proposed method can be used for detecting gradual transitions and abrupt changes without requiring any training of the video content in advance. Experiments have been conducted on a multi type video database and show that this algorithm achieves well performances.

Copyright © 2017 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:
Zaynab El khattabi
Faculty of Sciences, Abdelmalek Essaadi
Tetuan, Morocco
zaynabelkhattabi@gmail.com

1. INTRODUCTION

The high increasing volume of video content on the Web has created profound challenges for developing efficient indexing and search techniques to manage video data. Whereas managing multimedia data requires more than collecting the data into storage archives and delivering it via networks to homes or offices, content based video retrieval is becoming a highly recommended trend in many video retrieval systems. However, conventional techniques such as video compression and summarization strive for the two commonly conflicting goals of low storage and high visual and semantic fidelity [1].

Video segmentation is the fundamental process for a number of applications related to automatic video indexing, browsing and video analysis. The basic requirement of video segmentation is to partition a video into shots. It is often used as a basic meaningful unit in a video. In [2], Thompson et al. defined a video shot as the smallest unit of visual information captured at one time by a camera that shows a certain action or event. Therefore, segmenting video into separate video shots needs to detect the joining of two shots in the video and locate the position of these joins.

There are a number of different types of transitions or boundaries between shots. A cut is an abrupt shot change that occurs in a single frame. A fade is a slow change in brightness usually resulting in or starting with a solid black frame. A dissolve occurs when the images of the first shot get dimmer and the images of the second shot get brighter, with frames within the transition showing one image superimposed on the other. A wipe occurs when pixels from the second shot replace those of the first shot in a regular pattern such as in a line from the left edge of the frames [3]. Other types of shot transitions include computer generated effects such as morphing. The effects of this kind of transition are obtained with the help of the cross-dissolve or fading techniques which permit to achieve a smooth change of image content (i.e. texture and/or color) from source to target frames.

Whereas there is a wealth of research on shot boundary detection (SBD), some methods aim at detecting

abrupt boundaries, while others focus on gradual boundaries. In addition, certain kind of transitions can be easily confused with camera motion or object motion.
In this paper, a shot boundary detection scheme based on SIFT is proposed. Section 2. presents the various methods that have been proposed in this field, section 3. presents the method. Finally, section 4. and 5. give the experiments and a conclusion.

2. RELATED WORKS

In literature, Algorithms for shot boundary detection can broadly be classified into many groups; we can find lots of techniques include comparison of pixel values, statistical differences, histogram comparisons, edge differences, compression differences, and motion vectors to quantify the variation of continuous video frames.
The easiest way to detect if two frames are significantly different is to count the number of pixels that change in value more than some threshold. This total is compared against a second threshold to determine if a shot boundary has been found. Only the luminance channel of the considered videos is considered in this case. If the number of pixels which change from one image to another exceeds a certain threshold a shot transition is declared [4]. A technique introduced and validated during the TRECVID 2004 campaign is presented in [5]. First, small images are created from the original frames by taking one pixel every eight pixels and they are converted to HSV color space, only the V component is kept for luminance processing. With every new frame, the absolute difference between pixels intensity is computed and compared with the average values to detect cut transitions. Regarding the gradual transitions the method can detect only dissolves and fades. The idea proposed in [6] is dividing the images into 12 regions and founding the best match for each region in a neighborhood around the region in the other image. Gradual transitions were detected by generating a cumulative difference measure from consecutive values of the image differences.The inconvenient of methods based on comparison of pixel values is their sensitivity to camera motion.

To avoid this problem of camera motion and object movements, some techniques can be done by comparing the histograms of successive images. The idea behind histogram-based approaches ([7], [8]) is that two frames with unchanging background and unchanging (although moving) objects will have little difference in their histograms. Color histograms are used in [9] to detect shot boundaries by representing each frame of the video by their color histogram features. Then, the video frames are treated as a sequence of feature vectors which are fed to the split and merge framework. After completion of recursive split and merge process, the shot boundaries are identified easily.

Another approach to detect shot boundaries is edge/contour-based methods that exploit the contour information present in the individual frames, under the assumption that the amount and location of edges between consecutive frames should not change drastically. In [10], the feature of edge pixel count is proposed for shot detection, where Sobel edge detector is used. Besides, color, edge or texture information can be combined to make use of the advantages of all this features and increase the accuracy of the technique used. An example of this combination is proposed in [11] using global color features combined with the characteristics of local edge.

Some temporal filtering mechanism is used to eliminate camera motion noise when it is present in detecting shot changes. The work analysis resides in the discrimination between camera work-induced apparent motion and object motion-induced apparent motion, followed by analysis of the camera work-induced motion in order to identify camera work [12]. In [13], an approach block-based motion estimation is used, in which the whole frame is divided into possible blocks of 3x3 pixels. All pixels within the same block are assumed to belong to the same object, which undergoes translational motion. Each block is compared with all possible such blocks within the corresponding search window with the same center pixel location in current frame. In an other side, a camera motion characterization technique is introduced in [14] using a camera motion histogram descriptor to represent the overall motion activity of a shot.

Various features can be combined to make use of the advantages of various popular techniques such as color, texture, shape and motion vectors in spatial as well as in transformed domains such as Fourier, cosine wavelets, Eigen values, etc. An example of such combinations is presented in [15] where color feature is used and in [16], where texture feature is used. Texture methods like Local Binary Patterns (LBP) are used in various recent computer vision and pattern recognition applications. In [16] an extension of LBP histogram is used to represent the frame texture, it is called Midrange LBP (MRLBP). The authors justify their proposition by the comparison of gray center pixel value, average gray value and midrange gray value that is more robust to noise and illumination variants. LBP histogram values are extracted based on midrange statistics on each frame and they are stored as a feature vector in a video sequence. Then, the dissimilarity metric is applied on the feature vectors of adjacent frames to be used for shot detection process using adaptive threshold approach.

Shot boundary detection approaches can also be categorized based on machine learning techniques such as support vector machines, neural networks, fuzzy logic, clustering techniques and Eigen analysis [17] . In this context, the problem of shot detection in endoscopic surgery videos is addressed in [18] to manage the video content of surgical procedures. The method proposed relies on the application of a variational Bayesian (VB) framework for computing the posterior distribution of spatiotemporal Gaussian mixture models (GMMs). The video is first decomposed into a series of consecutive clips of fixed duration. Then, the VBGMM algorithm is applied on feature vectors extracted from each clip to handle automatically the number of components which are matched along the video sequence. These components denote clusters of pixels in the video clip with similar feature values and the labels are the tags of these components. Hence, the process of label tracking starts to define shot borders when component tracking fails, signifying a different visual appearance of the surgical scene. Genetic Algorithm and Fuzzy Logic have been also used for shot boundary detection. The authors of [19] proposed a system based on computing the Normalized Color Histogram Difference between each two consecutive frames in a video. Then, a fuzzy system is performed to classify the frames into abrupt and gradual changes. In order to optimize the fuzzy system, genetic algorithm GA is used. The results show the benefits of the GA optimization process on achieving a low computational time.

Many recent approaches reported in the literature related to shot boundary detection rely on SIFT ([20], [21]). The method proposed in [20] is based on SIFT-point distribution histogram extraction. Each video frame is represented by a histogram, named SIFT-point distribution histogram (SIFT-PDH). It describes the distribution of the extracted stable keypoints within the frame under polar coordinates. Distance comparison represents the difference between each two consecutive frames of the video; it is calculated by comparing their SIFT-PDHs. An adaptive threshold is used to identify the shot boundaries. Some other surveys of existing SBD techniques in the literature are provided and discussed in [22].

3. PROPOSED METHOD

Selection of an appropriate approach feature for segmenting a video sequence into shots is the most critical issues. Several such features have been suggested in the literature (histogram difference, optical flow...), but none of them is general enough to operate for all of changes in the video data.

The proposed method is based on feature extraction using scale invariant feature transform adopted by David G. Lowe [23]. The reason of this choice is that the SIFT image features are invariant to image rotation, scale and robust across a substantial range of affine distortion, addition of noise, and change in illumination. Firstly, the video is overviewed and zooms in wherever a shot boundary exists using a top down search scheme that is presented in [24]. The search is carried out by comparing the ratio of matched keypoints extracted via SIFT for every RGB channel of two video frames separated by a temporal sampling period N. SIFT descriptors are computed over all three channels of the RGB color space. Hence, three feature descriptors matrices associated with R, G and B color spaces have been obtained for each N^{th} frame. Instead of comparing the number of SIFT feature key points, we calculate and compare the ratio of matched number to total number between every two sampled frames to avoid false detection caused by too few keypoints generated. In order to zoom into the location of boundaries, peaks are detected and filtered to take only the deep enough peaks to be regarded as boundaries.

3.1. Feature Extraction

Scale Invariant Feature Transform (SIFT) is an approach for detecting and extracting local feature descriptors that are reasonably invariant to changes in illumination, image noise, rotation, scaling, and small changes in viewpoint. There are four major steps: Detection of scale-space extreme, accurate keypoint localization, orientation assignment, descriptor representation.

- **scale-space peak selection:** The first stage of computation searches over all scales and image locations. It is implemented efficiently by using a difference-of-Gaussian function (DoG) to identify keypoint candidates for SIFT features that are invariant to scale and orientation. DoG scale space can be obtained from equation (1).

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \tag{1}$$

where * is the convolution operation,*I(x,y)* is the gray value of pixel at *(x,y)* and *G(x,y,σ)* is a variable-scale Gaussian kernel defined as:

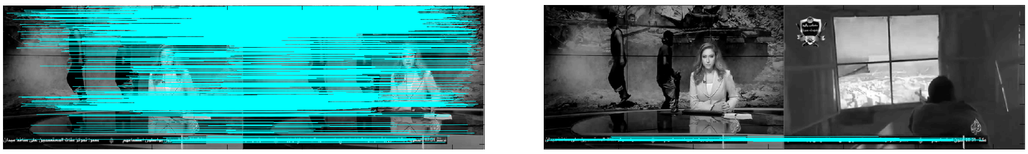
$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \tag{2}$$

- **Keypoint localization:** At each candidate location, a detailed model is fit to determine location and scale. Keypoints are selected based on measures of their stability. Low contrast keypoints introduced by noise and edge response will be removed.
- **Orientation assignment:** An orientation is assigned to each keypoint to achieve invariance to image rotation. A neigbourhood is taken around the keypoint location depending on the scale, and the gradient magnitude and direction is calculated in that region. An orientation histogram with 36 bins covering 360 degrees is created.
- **keypoint descriptor:** A 16x16 neighborhood around the keypoint is taken. It is divided into 16 sub-blocks of 4x4 sizes. For each sub-block, 8 bin orientation histogram is created. So a total of 128 bin values are available. It leads to a SIFT feature vector of 128 dimensions.

Color provides more discriminatory information than simple intensities. Although, RGB Color space is simple and very common. Hence, in our work, SIFT descriptors are computed for every RGB channel independently, and the information available in the three different color spaces are combined, unlike SIFT model that is designed only for grayscale information and misses important visual information regarding color.

3.2. Shot boundary detection

SIFT keypoints are extracted from frames of video and then ratios of matched keypoints number to total number between frame i and frame *i+N* are used to detect shot boundaries. The advantage of feature matching is that it is invariant to affine transformations; thus, we can even match objects after they have moved. Figure 1 shows local feature matching between two frames.



(a) Frames within the same shot. (b) Frames from different shots.

Figure 1. Feature keypoints matching between two frames.

The similarity matching between two frames in the same shot is usually high, due to the similar image feature, objects and colors. However, frames from different shots have visual discontinuity. As a result, they have no similarity matching or a low number of it.

3.2.1. The top down search scheme

To avoid unnecessary processing of video frames within any shot, a search is first carried out by performing similarity matching for every N^{th} frame in the video. It is a good solution for decreasing computational cost. Let us denote the i^{th} frame of a video as *F(i)*. Then, the algorithm is conducted as follows (Figure 2):

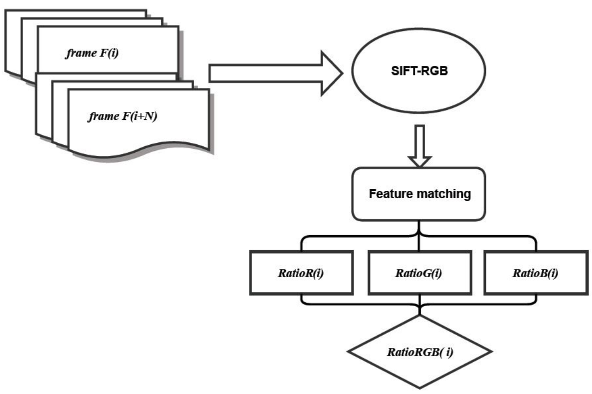


Figure 2. The top down search process.

Each color channel obtained for each N^{th} frame of the video is subjected to feature extraction process (SIFT-RGB), the output of which is fed to similarity matching process among the successive frames that results in three similarity values for each *i* frame: ratioR, ratioG and ratioB. This similarity information is fused to obtain one ratio representing the matched similarities between *F(i)* and *F(i+N)*.
The choice of using the ratios of matched features extracted to total number features, instead of comparing the number of feature keypoints with a prefixed threshold, is referred to the false detection caused by the small number of keypoints in the frames with few objects and colors, which generates a fewer matched similarities even though they are similar. The ratio for each color channel of the frame F_i is defined as:

$$ratioR(i) = \frac{2M_r}{K_r(F_i) + K_r(F_{i+N}))} \tag{3}$$

$$ratioG(i) = \frac{2M_g}{K_g(F_i) + K_g(F_{i+N}))} \tag{4}$$

$$ratioB(i) = \frac{2M_b}{K_b(F_i) + K_b(F_{i+N}))} \tag{5}$$

Where M_r,M_g and M_b are the number of matches found respectively for red, green and blue color planes between F_i and F_{i+N}. K_r,K_g and K_b are the total number of feature keypoints extracted from each color plane of the frame .The final ratio obtained from the three ratios is defined as:

$$Ratio_{RGB}(i) = \frac{ratioR + ratioG + ratioB}{3} \tag{6}$$

The determination of the temporal sampling period *N* depends on the type of video content and the duration of the shots, if a sequence of successive frames is captured by many cameras like in case of action movies, we can have uncontinuous action and very short shots. Consequently, an entire shot may start and end up between the sampled frames and be missed. For that, the choice of *N* must take into consideration the nature of video content. The temporal sampling period *N* is chosen to be *N=25 (1 sec)* in the example illustrated in figure 3.

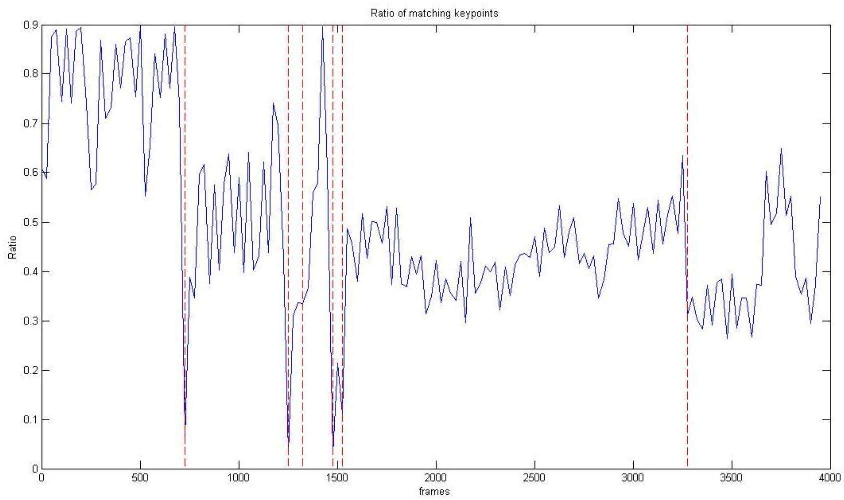


Figure 3. the overview of a video with N=25.

In order to zoom into locations of shot boundaries, extrema peaks are detected to filter the very deep peaks to be taken as boundaries. The peak detection function is used in [24] to find boundaries by comparing each minima peak with the previous and successive extrema peaks, using a threshold *T=0.5* to compare the depth of the peak with the others. The boundaries detection function is described in Algorithm 1.

P_i is a peak and P_t and P_r are the left and right end of the peak. Dashed lines in figure 3 present the peaks detected with this function.

Algorithm 1:Boundaries detection

For i=1,2,3,... **do**
  if ($P_i < P_{i-1}$ and $P_i < P_{i+1}$)
    then t=i-1; r=i+1;
      while ($P_t < P_{t-1}$) t=t-1;
      while ($P_r < P_{r+1}$) r=r+1;
      if($P_i < P_t$*T or $P_i < P_r$*T)
        then zoom in to [$F_{(i-1)*N}$,F_{i*N}]

3.2.2. Determination of transition type

To determine if a shot is a hard cut or gradual transition, the moving average value of frames in the boundaries is calculated. The moving average of frame *t* is defined as:

$$AverageRatio(t) = \frac{1}{N}\sum_{i=t-N}^{t-1} Ratio_{RGB}(t) \tag{7}$$

Where $Ratio_{RGB}(t)$ is the ratio of matching feature keypoints obtained in equation (6) by fusing the three ratios *ratioR*, *ratioG* and *ratioB* of a frame *t*, this frame is detected as a boundary using the algorithm 1. The period N is used as a number of previous frames used with the current frame *t* when calculating the moving average. We can distinguish transitions by measuring the difference of *AverageRatio(t)* and $Ratio_{RGB}(t)$ as described in algorithm 2.

Algorithm 2: Type of transition

For $t = t_1, t_2, ..., t_n$ **do** (t_i is a shot boundary)
  if ($AverageRatio(t) - Ratio_{RGB}(t) >= \alpha$)
  then
   type of transition=cut boundary
  else
   type of transition=gradual transition

A threshold α is used to detect transition types. In our experiments, the choice of an appropriate threshold α, has a high impact on the accuracy of the results.

4. EXPERIMENTS AND RESULTS

In order to evaluate the performance of the proposed method and reveal its advantages over the other methods in literature, We have designed an experimental video dataset containing four types of videos (sport, news, cartoon, movie) .The video sequences used are MPEG-4 compressed videos, with various dimensions and containing several types of transitions, The Experiment dataset used for evaluation are listed in table 1.

Table 1. Information of experimental videos

Type	Number of frames	Size	Duration	Number of shots
Sport	83525	640x360	3341 sec	411
News	45100	640x360	1804 sec	223
Cartoon	31855	1280x720	1385 sec	204
Movie	72749	1280x720	3163 sec	530

The performance results of the proposed method are shown as precision and recall values in Table 2. Precision and recall are defined as:

$$Precision = \frac{N_c}{N_c + N_f} \tag{8}$$

$$Recall = \frac{N_c}{N_m + N_c} \tag{9}$$

Where N_c,N_f and N_m are the numbers of correct, false and miss shot boundary detections, respectively.

Table 2. Evaluation of the proposed method

	Abrupt Changes		Gradual Transition	
	Precision	Recall	Precision	Recall
Sport	0.92	0.85	0.93	0.77
News	0.95	0.94	0.89	0.86
Cartoon	0.88	0.91	0.75	0.81
Movie	0.94	0.87	0.79	0.88

Figure 4 shows some shot boundaries detected from the experimental dataset. The transitions presented in figure 4 belong to a cut transition where there is a complete dissimilarity between two successive frames, and the ratio of matched keypoints is very small or null.

(a) Example 1 of cut transition (frames 99 and 100).

(b) Example 2 of cut transition (frames 230 and 231).

Figure 4. Examples of two cut transitions detected in cartoon video.

We tested our method on some videos from the Open Video Project [25]. Figure 5 shows the frames in the first gradual transition detected by our method on a video provided by The Open Video repository: (NASA 25th Anniversary Show, segment 1), we can see clearly that the changes and dissimilarities occur gradually between the successive frames. These variations are translated by the value of RGB ratio of matched similarities that decrease gradually between the frame 128 and 142.

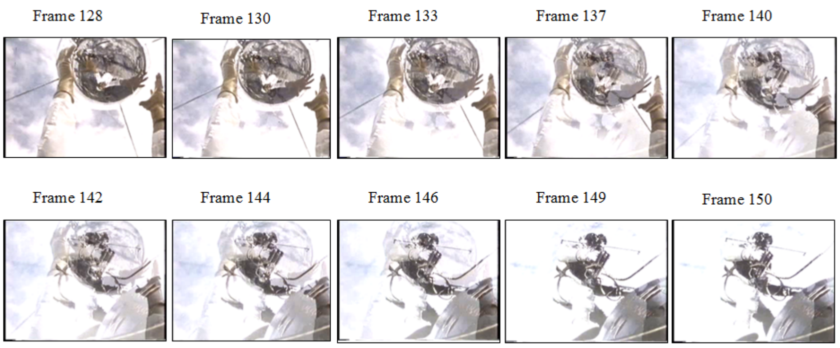


Figure 5. Example of gradual transition detected.

The low recall rate in sports video is may be due to the short shots that are missed between the sampled frames. In contrast, the precision rates in this kind of videos are more than 90%. It shows that the method is effective in detecting abrupt and gradual transitions.On the other side, in general, recall rates are low. This reveals that some frames belonging to different shots were regarded as similar. As a result, several shot boundaries are missed.

In news video the precision rate and the recall rate are high (more than 90 %),because of the long shots and the existence of many cut transition which are distinguished by the great changes between the frames. Accordingly, shot boundaries are well detected. Also, the choice of the temporal sampling period N as 1 second indicates that all the shots less than this value will be missed. The adaptation of the parameter N in accordance with the video sequences can increase the performance results by the reduction of miss or false shot boundary detection.

The comparison of this method with the experimental results reported in other works based on SIFT, shows that the integration of the three color channels R, G and B of video frames gives more precision in detecting shot boundaries than using only the grayscale channel.

5. CONCLUSION

In this work, a new algorithm is presented based on scale invariant feature transform adapted to the RGB color space. First, a top down search process is performed by comparing the ratio of matched keypoints extracted via SIFT for every R, G and B channels of two video frames separated by a temporal sampling period N. Then, an algorithm is used to detect the shot boundaries. Finally, the moving average of frames in the boundaries is calculated to determine the type of the transition by using a threshold. Our method is applied to different types of video and shows satisfactory performance in detecting abrupt changes and gradual transitions, but it can be improved by using weighting coefficients to calculate the ratioRGB from the three ratios(R,G and B), depending on the type of the video. In the future works,we aim to include performance improvements and minimizing the computational cost without decreasing the accuracy.

REFERENCES

[1] J. T. T. Mei, L.-X. Tang and X.-S. Hua, "Near-lossless semantic video summarization and its applications to video analysis," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 9, no. 3, June 2013.

[2] R. Thompson, *Grammar of the Shot*, F. Press, Ed., 1998.

[3] J. S. Boreczky and L. A. Rowe, "Comparison of video shot boundary detection techniques," *Journal of Electronic Imaging*, vol. 5, no. 2, pp. 122–128, April 1996.

[4] R. G. Tapu, "Segmentation and structuring of video documents for indexing applications," December 2012.

[5] S. H. G. Jaffre, Ph. Joly, "The samova shot boundary detection for trecvid evaluation 2004," in *Proceedings of the TRECVID 2004 Workshop, Gaithersburg, MD, USA, NIST*, 2004.

[6] B. Shahraray, "Scene change detection and content-based sampling of video sequences," in *Proc. SPIE Digital Video Compression: Algorithms and Technologies*, vol. 2419, 1995, pp. 2–13.

[7] C.-L. Huang and B.-Y. Liao, "A robust scene-change detection method for video segmentation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 12, pp. 1281–1288, December 2001.

[8] D. S. Guru and M. Suhil, "Histogram based split and merge framework for shot boundary detection," *Mining Intelligence and Knowledge Exploration, Lecture Notes in Computer Science*, vol. 8284, pp. 180–191, December 2013.

[9] D. Guru and M. Suhil, "Histogram based split and merge framework for shot boundary detection," *Mining Intelligence and Knowledge Exploration, Lecture Notes in Computer Science*, vol. 8284, pp. 180–191, December 2013.

[10] S. C. R. S. Jadon and K. K. Biswas, "A fuzzy theoretic approach for video segmentation using syntactic features," *Pattern Recognition Letters*, vol. 22, no. 13, pp. 1359–1369, November 2001.

[11] L. Y. R. L. C. Y. . Z. R. Qu, Z., "A method of shot detection based on color and edge features," in *1st IEEE Symposium on Web Society, SWS'09*, August 2009, pp. 1–4.

[12] H. Z. P. Aigrain and D. Petkovic, "Content-based representation and retrieval of visual media: A state-of-the-art review," *Multimedia Tools and Applications*, vol. 3, no. 3, pp. 179–202, November 1996.

[13] S. M. P. Panchal and N. Patel, "Scene detection and retrieval of video using motion vector and occurrence rate of shot boundaries," in *2012 Nirma University International Conference on Engineering (NUiCONE)*, December 2012, pp. 1–6.

[14] X. H. Y. W. Muhammad Abul Hasan, Min Xu, "A camera motion histogram descriptor for video shot classification," *Multimedia Tools and Applications*, vol. 24, no. 74, p. 1107311098, December 2015.

[15] F. B. F. Bayat, M. Shahram Moin, "Goal detection in soccer video: Role-based events detection approach," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 4, no. 6, pp. 979–988, 2014.

[16] . N. H. S. Rashmi, B. S., "Video shot boundary detection using midrange local binary pattern," in *International Conference on Advances in Computing, Communications and Informatics (ICACCI),IEEE*, September 2016, pp. 201–206.

[17] A. M. E. M. M. Pournazari, F. Mahmoudi, "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.

[18] N. N. S. D. . G. E. Loukas, C., "Shot boundary detection in endoscopic surgery videos using a variational bayesian framework," *International journal of computer assisted radiology and surgery*, vol. 11, no. 11, pp. 1937–1949, 2016.

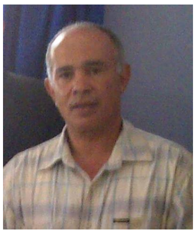
[19] K. T. S. K. M. . R. S. Thounaojam, D. M., "A genetic algorithm and fuzzy logic approach for video shot boundary detection," *Computational intelligence and neuroscience*, no. 14, 2016.

[20] E. A. A. K. N. P. . J. M. Hannane, R., "An efficient method for video shot boundary detection and keyframe extraction using sift-point distribution histogram," *International Journal of Multimedia Information Retrieval*, vol. 2, no. 5, pp. 89–104, 2016.
[21] W. X. Z. W. . H. P. Liu, G., "Shot boundary detection and keyframe extraction based on scale invariant feature transform," in *Eighth IEEE/ACIS International Conference on Computer and Information Science, ICIS 2009*, June 2009, pp. 1126–1130.
[22] W. X. Z. A. . W. J. Chi, A., "Review of research on shot boundary detection algorithm of the compressed video domain in content-based video retrieval technique," in *DEStech Transactions on Engineering and Technology Research, (iceta)*, 2016.
[23] D. Lowe, "Distinctive image features from scale invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
[24] M. Birinci and S. Kiranyaz, "A perceptual scheme for fully automatic video shot boundary detection," *Signal Processing: Image Communication*, vol. 29, no. 3, pp. 410–423, March 2014.
[25] The open video project. [Online]. Available: https://open-video.org/index.php

BIOGRAPHIES OF AUTHORS

Zaynab El khattabi is a Ph.D. student in Faculty of Sciences, Abdelmalek Essadi University, Morocco . She is a Computer Sciences engineer, graduated in 2012 from National School of Applied Sciences, Abdelmalek Essadi University. She got a DEUG on Mathematics and computer Sciences in 2009 from Faculty of Sciences, Abdelmalek Essadi University. Her current research interests include image and video processing and focuses on video-content analysis and retrieval.

Youness Tabii received his PhD in July 2010 from the National School of Computer Sciences and Systems Analysis, Mohammed V University-Rabat. He is a Professor at the National School of Applied Sciences of Tetuan (ENSAT). He is a member in New Technology Trends Team (NTT Team) and the Head of Master: Embedded and Mobile Systems. His research interest includes video processing and analysis, also interested by cloud security. He is the Founder and Chair of International Conference on Big Data, Could and Applications (BDCA). He is a Guest-Editor of the International Journal of Cloud Computing in 2016.



Abdelhamid Benkaddour got a MAS and a PhD in Applied Mathematics and Mechanics from Pierre et Marie Curie (Paris VI) University in June 1986 and 1990, respectively, and a PhD in Mathematics from Abdelmalek Essaadi University in 1994. His research focuses on numerical analysis, scientific computing and computer science.