# Face Detection in Close-up Shot Video Events Using Video Mining

Amjad Rehman Khan [1], Majid Harouni [2,*], Sepideh Sharifi [2], Saeed Ali Bahaj [3], and Tanzila Saba [1]

[1] Artificial Intelligence & Data Analytics Lab CCIS, Prince Sultan University Riyadh 11586, Saudi Arabia; Email: {arkhan, tsaba}@psu.edu.sa (A.R.K., T.S.)

[2] Department of Computer Engineering, Dolatabad Branch, Islamic Azad University, Isfahan, Iran; Email: sepideh.sh71@yahoo.com (S.S.)

[3] MIS Department, College of Business Administration, Prince Sattam bin Abdulaziz University, Alkharj 11942, Saudi Arabia; Email: s.bahaj@psau.edu.sa (S.A.B.)

*Correspondence: majid.harouni@gmail.com (M.H.)

***Abstract*—Face detection and recognition in abrupt dynamic images is still challenging due to high complexity of images. To tackle this issue, we employed Gray-Level Co-occurrence Matrix (GLCM) to convert large video into smaller consequential sections containing sequence information from a series of images. GLCM is a matrix associated with the relationship between the values of adjacent pixels in an image. The proposed method is composed of two stages. First, the video is taken as input using the histogram difference method. Features are extracted using co-occurrence matrix of images, statistical methods, and the border of sudden shots extracted from the video. Second, face recognition with the Viola-Jones algorithm is performed on the sudden shots extracted in the first step. Thus, the face is extracted by video data mining in output in close shots. In this method, we compared the parameter model in three windows (3, 5 and 7) and threshold limit for detecting abrupt cuts between values (0.1, 0.5, 1.5, 1.5 and 2) for each window. The highest percentage of face detection is attained by considering the maximum percentage of abrupt cuts in the 5×5 window with a threshold value of 1.**

***Keywords*—close shot, statistical methods, co-occurrence matrix of images, face detection, security, technological development**

## I. INTRODUCTION

Due to the huge volume of stored video data, the need for strategies to explore such big data has been considered more than ever in the past few years. Storing, managing, retrieving and analyzing video data imposes heavy costs on the process in terms of time and algorithms. In other words, video data exploration involves combining methods of analyzing and retrieving this type of data and performing exploration methods. Also, the continuous rise in the data volume is another challenge. Among multimedia data, video is one of the most widely used examples [1, 2]. Organizations, institutions, and individuals generate and disseminate such data daily. The widespread use of this type of multimedia data have been seen in all military, security, scientific and other fields, has caused researchers and scholars to pay special attention to it [3, 4]. Unlike other examples of data types, exploring video data is a relatively new and evolving field of work. Hence, there is a need of tools to discover the connections between objects or pieces in the video, such as framing video images based on their content or extracting patterns from them. However, it is associated with high complexity [5, 6]. Discovering duplicate image patterns is essential to reduce the size of the data dimension, this will make it easier to understand, organize and search the image content. Dividing video into smaller units is one of the most essential and basic steps of a video data management system, which is one of the main challenges [7, 8]. This allows to process video shots and specify frames that show a face or non-face. In this paper, face detection and recognition challenge has been investigated in close shots.

Face detection is now one of the most important research areas that has been considered by many researchers and its applications include security control of individuals, access control of criminals and reconstruction of the face [9]. Facial features significantly impact human interactions and show an immediate sense of the individual's moods [10]. However, face recognition accuracy in different moods with various expressions still needs to improve [11, 12].

Accordingly, this research has the following main contributions.

- Proposed a framework to detect and classify faces in close abrupt images with various expressions.
- Achieved high accuracy for locating faces in close abrupt images with complex background.

Further, this research has four main sections, Section II explores research background, Section III presents proposed methodology, Section IV exhibits experimental results and finally Section V concludes research.

## II. BACKGROUND

In video data mining, it is important to discover duplicate image patterns to reduce the size of the data dimensions [13]. This will make it easier to understand,

Manuscript received August 10, 2022; revised September 12, 2022; accepted September 28, 2022; published March 8, 2023.

organize and search the image content. A histogram-based border detection presented using CBP index, in which the comparison of consecutive frame histograms was used to reveal the shot edge [14]. Mishr *et al.* [15] proposed a method for detecting video shot transmission based on wavelet transform and wavelet entropy, the mean and standard deviation in the low frequency of an image used to detect the cut transfer [16]. Liu [17] presents a new method for recognizing shot boundaries in a video stream; in this method, first the video streams are divided into different frames, then on each frame, processing processes have been performed to extract the features, and the maximum average local vector of the external data of each frame is considered as the boundary of a shot [18]. Finally, an adaptive two-threshold algorithm is proposed for video game playing on TV. In this algorithm, each frame is first transformed to HSV space; the upper and lower threshold ratios are adjusted according to the difference in frames. Then the average value of the frame difference is calculated by identifying the two thresholds [19]. Most of the state-of-the-art research described a block matching algorithm and complex wavelet transform based on comparing two detection techniques using various video sequences.

Another study developed a technique for recognizing a video shot boundary using a Gray Level Co-occurrence Matrix (GLCM) to locate a rapid transition between two successive images. First, each frame is converted to grayscale, then the GLCM-based correlation coefficient is determined from successive video frames, and the video's border is threshold [20, 21]. Data mining involves extracting hidden information, patterns, and correlations from enormous amounts of data [22, 23]. To minimize data dimensionality in video data mining, redundant visual patterns must be found.

Liu *et al.* [17] organized video frames using theoretical graphical techniques, and block displacement was employed to identify shot changes. Displaced Frame Difference (DFD) and flexible planar motion similarity are used to identify shots. Another research developed a method that retrieved structural information from each video frame using a wavelet transform binary tree [18], then assessed the structural similarity of succeeding frames' geographical range [19]. On various video sequences, a comparison was made between the SBD block matching technique and the binary tree-based SBD algorithm of complex wavelet transform. A video border identification system based on segment selection and SVD with pattern matching.

Pal *et al.* [20] determined the shot boundaries and gradual transition duration using adaptive thresholds, and most non-boundary frames are deleted concurrently. Research examined shot boundary identification using block-based characteristics. This measurement involves not halting the camera and moving the object/background to identify the shot based on Displaced Frame Difference (DFD) and block-like motion [21]. A different effort used multi-fractal analysis to recognize video boundaries. Low-level features (colour and texture) are taken from each video frame and connected to Feature Vectors (FVs) in the feature matrix. The matrix row corresponds to the frame feature vector, while the columns are the unique feature vector components [23]. A model based on shot boundary detection utilizing frame transfer parameters estimates the frame using the previous and next frames [4]. A keyframe-based video summary employing adaptive edge rates and auto thresholds is suggested. First, the histogram difference of each frame is determined, then the Prewitt operator extracts the important frame edges [24]. Video shot boundary identification utilizing complex binary tree wavelet transform was presented for decoding encoded video sequences. The suggested approach initially extracts video frame structure using dual-tree wavelet transform. Next, structural similarity of the spatial domain across frames is calculated [25]. A rapid SVD-based shot boundary identification and pattern matching is described (SVD). The shot borders and transition duration are predicted using compatibility criteria, and most non-boundary frames are eliminated.

A video summary employing shot detection is suggested and examined; this analysis uses the difference between histogram and block based on changing Euclidean distances [26]. In another research, face recognition methods such as Viola-Jones, SMQT features, SNOW classification, neural network-based face identification, and machine-based face vector support were examined. These face recognition systems are evaluated based on accuracy and remaining quantity using DetEval [27]. The Viola-Jones known as a really effective method in face recognition.

Al-Tuwaijari *et al.* [28] applied many algorithms for feature extraction (LDA algorithm) and feature reduction (BAT and chicken algorithms), classification (J48 algorithms) as well as viola jones for main face detection algorithm.

## III. Proposed Method

The primary purpose of this research is to design and implement an optimal face recognition method in close shots of video events with high data volume and minimal knowledge. In this method, we first reduce the volume of data by recognizing the shot boundary in the first step due to the high volume of video data. In the second step, with the Viola-Jones algorithm, we recognize the face in two stages of training and discovery. The proposed framework is exhibited in Fig. 1, composed of two main steps: recognizing the border of sudden shots and recognizing the face in close shots.

### A. Preprocessing for Each Frame

First, reading the video, in which an input video V is first to read and framed as V ($F_1$, $F_2$, …, $F_n$) and its general specifications are checked, then the video frames $F_1$, $F_2$, …, $F_n$ are read, respectively. Second, transformation to gray level, for this transformation, frames $F_1$, $F_2$, …, $F_n$ change from the RGB (red, green, and blue) color model to the gray level from Eq. (1):

$$G\_L(x,y)=0.30*R(x,y)+0.59*G(x,y)+0.11*B(x,y) \quad (1)$$

After that, calculating the histogram of each frame of the video, in the image histogram, we have a graph in which the number of pixels of the brightness level in the input image is specified, each of which is a value in the range of 0 to 255. Therefore, to calculate the histogram, we have an array of length 256, which is a feature vector of the image. In Eq. (2), a brief and useful representation of intensity levels has been expressed in a gray level scale image:

$$H_i = \sum_{i=0}^{l-1} r_i p(r_i) \tag{2}$$

where $r_i$ is the *i*-th level of gray level and $p(r_i)$ represents the probability of occurrence.

Finally, calculating of the histogram difference between two consecutive frames, at this stage, the difference between the histogram of the current frame and the previous frame is calculated as follows. Eq. (3) gives us the histogram difference.

$$H_{i,i-1} = \sqrt{\sum_{i=1}^{n} ((H_i - H_{i-1}) \times (H_i - H_{i-1}))} \tag{3}$$

where, $H_i$ represents the current frame's histogram value and denotes the previous frame's histogram value, *i* is the frame number and *n* is the last frame number, which are calculated for all frames.

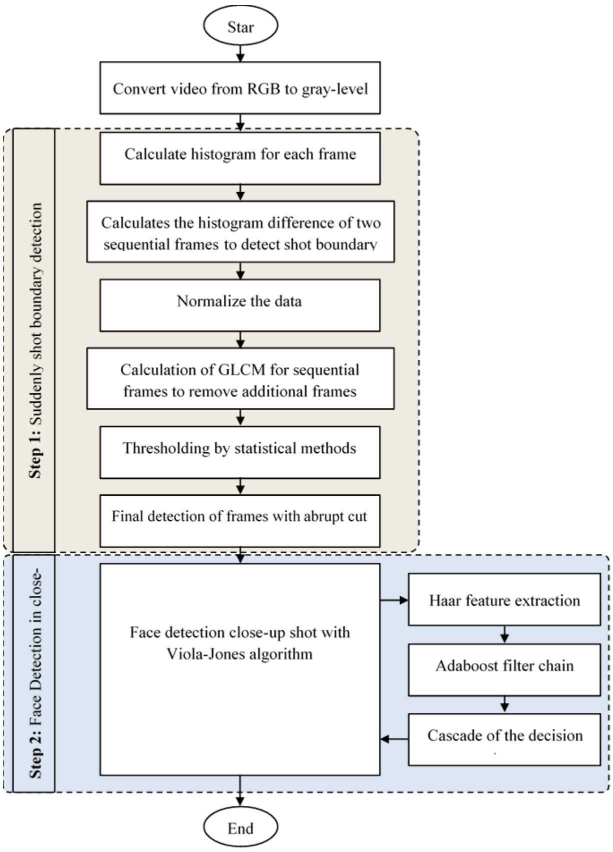


Figure 1. The proposed face detection framework.

## B. Extraction of the Features

*1) Detection of sudden shots:* After calculating the difference of the obtained histograms, by thresholding on them, we obtain abrupt cuts, for which we use the Dugad model. This model is used to calculate adaptive thresholds for each video. According to this model, the adaptive threshold significantly improves the detection speed regardless of the method used [28, 29]. Our parameters in this model include a threshold with a value of 3 and a window of 5×5. First, the right and left borders of the window are calculated. Then, it is checked that the obtained border is the maximum value of the histogram difference between the neighboring frames. If it has the highest value, it is considered an abrupt cut. The thresholds used have been stated in relationship in Eq. (4).

$$HD_s^{Condidate} = \underset{s\in N}{Max}\{HD_{s_{j-5}(x,y)} : HD_{s_{j+1}(x,y)} \mid \text{Where } i \text{ is} \{1,2,\ldots,k\} \text{ and } j \text{ is } K \bmod i = 0\} \tag{4}$$

$HD_s^{Candidate}$ represents the s-th candidate frame with the maximum histogram value in consecutive frames.

The detection of hard shots (candidate frames) has been done with the help of Eq. (5).

$$HD_s^{Candidate} > Thrd \begin{cases} \text{True} & HD_s^{Candidate} \text{ is candidate hard cut shot} \\ \text{False} & HD_s^{Candidate} \text{ is not candidate hard cut shot} \end{cases} \tag{5}$$

*2) Normalization of the value of candidate frame data*: With Eq. (6), normalization is performed between the output data between 0 and 1.

$$HD_{value}^{Candidate} = \frac{HD_{value}^{Candidate} - HD_{MINvalue}^{Candidate}}{HD_{MAXvalue}^{Candidate} - HD_{MINvalue}^{Candidate}} \tag{6}$$

*3) Calculation of GLCM for consecutive frames:* In this step for all frames, we calculate the GLCM for the current and next frames. The gray level matrix equals the number of rows and columns equal to the number of grey levels.

The element of the matrix *P* (*i*, *j* | Δx, Δy) is the number of repetitions of adjacent pixels in the whole image and the pixel distance (Δx, Δy), one with intensity i and the other with intensity j, occurring in a certain range. The matrix element *P* (*i*, *j* | Δx, Δy) may alternatively be considered to include second-order statistical probability values for changes between the grey levels i and j at a certain displacement distance d and at a particular angle (θ). Given the range M × N, for instance, an input picture may comprise grey levels G from 0 to G-1 and permit the intensity f (m, n) in sample m to be close to line n [30, 31].

$$P(i, j \mid \Delta x, \Delta y) = WQ(i, j \mid \Delta x, \Delta y)$$

$$W = 1 \big/ (M-\Delta x)(N-\Delta y)$$

$$Q(i, j \mid \Delta x, \Delta y) = \sum_{n=1}^{N-\Delta y} \sum_{m=1}^{M-\Delta x} A$$

and (7)

$$A = \begin{cases} 1 & \text{if } f(m,n)=i \text{ and } f(m+\Delta x, N+\Delta y)=j \\ 0 & \text{elsewhere} \end{cases}$$

The correlation coefficient $p_{xy}$ then examines the correlation between the two values and is calculated according to Eq. (8):

$$p_{xy} = \frac{n \sum_{p=1}^{n} \sum_{q=1}^{n} x_p y_q - \sum_{p-1}^{n} x_p \sum_{q=1}^{n} y_q}{\sigma_x \sigma_y} \tag{8}$$

$n$ represents elements' number in each frame; $\sigma_x$ and $\sigma_y$ are obtained from frames $x_p$ and $y_q$ according to Eqs. (9) and (10), respectively:

$$\sigma_x = \sqrt{n \sum_{p=1}^{n} x_p^2 - \left( \sum_{p=1}^{n} x_p \right)^2} \tag{9}$$

$$\sigma_y = \sqrt{n \sum_{q=1}^{n} y_q^2 - \left( \sum_{q=1}^{n} y_q \right)^2} \tag{10}$$

*4) Thresholding by statistical methods:* Our output is the number of frames and their correlation coefficient. We separate the more appropriate sudden by the mean and the median according to the following relationship.

$$x = 1 - p_{xy} \tag{11}$$

$$v_{mean} = 2 \times \left( \sum f(x) \big/ n \right) \tag{12}$$

$$v_{median} = \left(L + \left[\frac{\frac{n}{2} - F}{f_m}\right]\right) \times 2S \tag{13}$$

$$V = \left(\frac{\sum f(x)}{n}\right) + \left(\left(L + \left[\frac{\frac{n}{2} - F}{f_m}\right]\right) \times S\right) \tag{14}$$

$$V < 20 \begin{cases} \text{True} & \text{Vmean+Vmedian} \\ \text{False} & \text{Vmean} + 0.1 \end{cases} \tag{15}$$

where $p_{xy}$ represents the correlation coefficient, n number of frames, $L$ is the lower limit of the middle class, F is the cumulative frequency before the middle class and the absolute frequency of the middle class and $S$ is the distance of the classes.

Additional abrupt cuts are eliminated during these steps, and we have a suitable abrupt cut with a lower fault rate. We finally get the final abrupt cuts on the obtained values, placing 1 on the minimum value of the correlation coefficient and its range between the values of 100 to 10000.

*5) Face recognition based on the Viola-Jones algorithm:* After the above steps, from the resulting frames with abrupt cut, face recognition with Viola-Jones faces recognition algorithm so that the face range is detected. The lines are then drawn according to the four identified points for each component of the face and finally, we do the face detected in the output. Input parameters in this algorithm include the detection orientation of the face threshold, detection orientation of the threshold of the boundaries of the face components (right eye, left eye, nose and mouth) and the normalized face size, which we have set the face size to the default value of 176. Finally, we have output frames that are recognized if these are faces. An example of an output frame has been shown in Fig. 2.

Figure 2. Results of output frames.

## IV. Results and Analysis

To simulate the method presented in this research, we have used MATLAB and experimented database contains 158 videos, including movies and news, in which more faces are used.

### A. Dataset

Datasets provide standard training and testing images for experimental results analysis and comparison in state of the art [32]. The dataset experimented in this research includes shots extracted from video sequences of movies and news. Table I details the experimented database.

TABLE I. Database Properties

| | |
|---|---|
| Types | 5 movies - sports news |
| Number of movies | 158 |
| The total number of abrupt cuts (Hard Cut) | 496 |
| Time of each movie | Mac 20s |
| Movie format | .avi |
| Maximum processing time for each movie | Maximum: 0/319890869<br>Minimum: 0/002029328 |
| Frame rate | 24 f/s |

## *B. Performance Analysis*

Performance is a measure of precision and recall, precision means that among the shots with abrupt cut [33, 34]. How many images with abrupt cut we have in which the face is correctly recognized. The recall means that out of the number of shots in which the face has been correctly or incorrectly identified, how many abruptly cut images in which the face has been correctly identified have been expressed in Eq. (16) and Eq. (17).

$$\mathrm{Pr}esicion = TP / (TP + FP) \tag{16}$$

$$\mathrm{Re}call = TP / (TP + FN) \tag{17}$$

Table II presents the measurement parameters and their brief description.

TABLE II. ASSESSMENT PARAMETERS

| Descriptions | Assessment Parameter |
|---|---|
| Abruptly cut images in which the face has been correctly recognized. | TP |
| Abruptly cut images in which the face has not been recognized. | FP |
| Images are unrelated to abrupt cuts that are not recognizable and have a face. | TN |
| Images are unrelated to abrupt cuts that do not have a face and the face in them is misdiagnosed. | FN |

In Table III, the values of the measurement parameters have been calculated by changing the thresholds, which is the best threshold by calculating the threshold precision and recall values of 0.5 for windows 3 and 7, and the value of 1 is the best threshold in window 5. Similarly, Table IV and Table V present calculation of measurement criteria based on thresholding on the 5×5 window and 7×7 window.

TABLE III. CALCULATION OF MEASUREMENT CRITERIA BASED ON THRESHOLDING ON THE 3×3 WINDOW

| **Metrics** | **TRD=0/1** | **TRD=0/5** | **TRD=1** | **TRD=1/5** | **TRD=2** |
|---|---|---|---|---|---|
| TP | 113 | 119 | 113 | 84 | 46 |
| FP | 175 | 183 | 202 | 283 | 391 |
| TN | 197 | 197 | 195 | 163 | 119 |
| FN | 101 | 96 | 108 | 92 | 59 |

TABLE IV. CALCULATION OF MEASUREMENT CRITERIA BASED ON THRESHOLDING ON THE 5×5 WINDOW

| **Metrics** | **TRD=0/1** | **TRD=0/5** | **TRD=1** | **TRD=1/5** | **TRD=2** |
|---|---|---|---|---|---|
| TP | 110 | 113 | 116 | 85 | 47 |
| FP | 177 | 198 | 191 | 281 | 389 |
| TN | 200 | 198 | 200 | 157 | 122 |
| FN | 97 | 95 | 105 | 929 | 57 |

TABLE V. CALCULATION OF MEASUREMENT CRITERIA BASED ON THRESHOLDING ON THE 7×7 WINDOW

| **Metrics** | **TRD=0/1** | **TRD=0/5** | **TRD=1** | **TRD=1/5** | **TRD=2** |
|---|---|---|---|---|---|
| TP | 115 | 115 | 112 | 86 | 48 |
| FP | 179 | 185 | 210 | 272 | 387 |
| TN | 196 | 204 | 192 | 160 | 122 |
| FN | 94 | 91 | 98 | 90 | 58 |

## *C. Comparisons and Discussion*

The graphs have been plotted by examining the values of the variables detailed previously. However, we compared the proposed model in three windows (3, 5 and 7). In addition, the threshold limit for detecting abrupt cuts between values (0.1, 0.5, 1.5, 1.5 and 2) for each window was also compared. The results presented in Fig. 3 are of the 3×3 window with the mentioned thresholds.

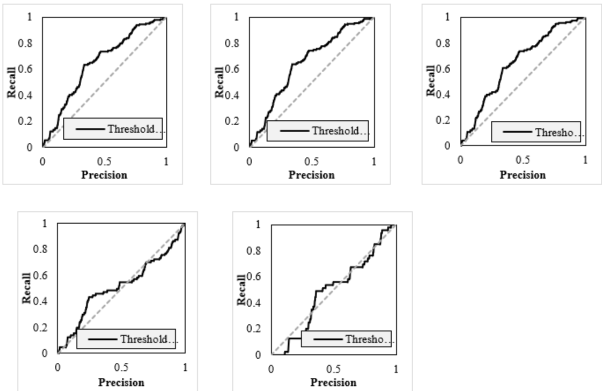

Figure 3. Results obtained for the 3×3 window with different threshold limit.

Fig. 4 presents 3×3 window, the threshold limit of 0.5 has the best efficiency and the following diagrams are the results of the 5×5 window with the mentioned thresholds. Therefore, the proposed method with window 5 and threshold 1 is the most optimal method.

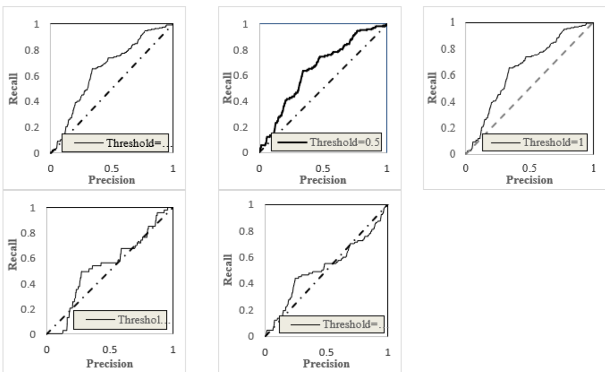

Figure 4. Results obtained for the 5×5 window with different threshold limits.

Finally, Fig. 5 shows results of the window with a value of 7 with the thresholds shown in the diagrams below. According to these diagrams, 0.5 also has the best efficiency, which, as can be seen, does not change the value of the optimal threshold as the window decreases or increases.

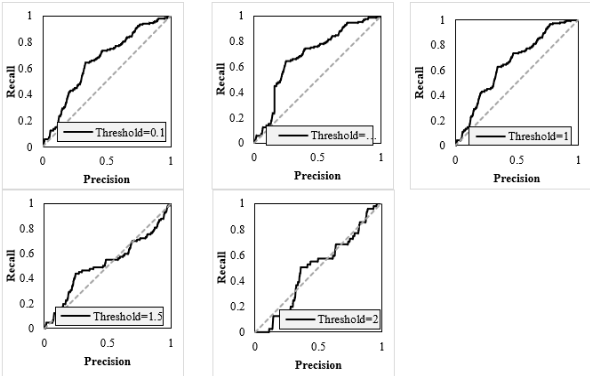


Figure 5. Results obtained for the 7×7 window with different threshold limits.

By observing and examining the above diagrams (Fig. 6) for each window, the threshold value of 0.5 in the 3×3 window, 1 in the 5×5 window and 0.5 in the 7×7 window provides us the best output that our proposed method with a 5×5 window and threshold value of 1 has the most efficient result among the studied methods. Table VI shows the highest percentage of face detection, considering the maximum percentage of abrupt cuts in the 5×5 window with a threshold value of 1, which is 34.66.

According to the results and review of windows 3 and 7, the number of additional abrupt and wrong cut values that we have in the output is high, which reduces the efficiency of work. However, we have the best in windows 5 with threshold 1.

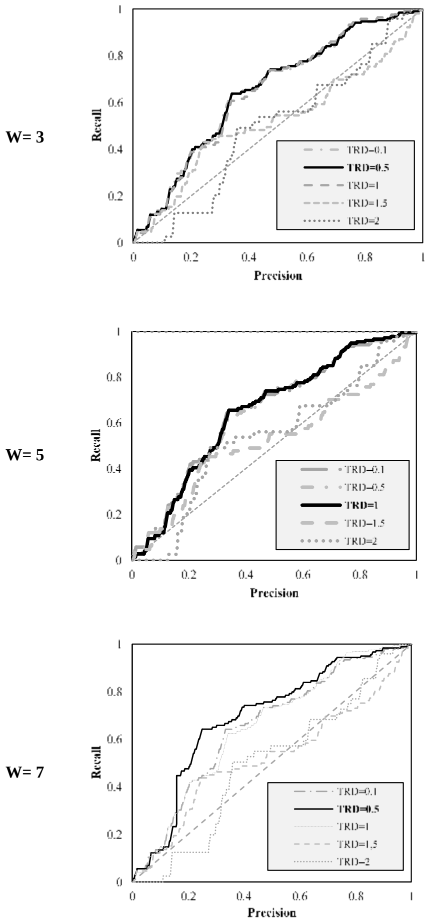


Figure 6. Presents a general diagram of the results from each window as follows.

TABLE VI. PERCENTAGE OF FACE RECOGNITION IN CLOSE SHOT

| Windows | Threshold Value | Number of output frame | Number of abrupt cuts | Number of faces in close shot | Number of face recognition in close shot | Percentage of abrupt cuts (%) | Percentage of face recognition in close shot (%) |
|---|---|---|---|---|---|---|---|
| 3×3 | 0/1 | 585 | 488 | 213 | 125 | 83/56 | 58/68 |
| | 0/5 | 595 | 474 | 218 | 138 | 79/66 | 63/30 |
| | 1 | 612 | 451 | 216 | 141 | 73/69 | 65/27 |
| | 1/5 | 753 | 406 | 312 | 201 | 53/91 | 64/42 |
| | 2 | 895 | 359 | 360 | 216 | 40/11 | 60 |
| 5×5 | 0/1 | 584 | 475 | 206 | 122 | 81/33 | 59/22 |
| | 0/5 | 595 | 468 | 214 | 132 | 78/65 | 61/68 |
| | 1 | 615 | 486 | 208 | 138 | 79/02 | 66/34 |
| | 1/5 | 753 | 407 | 314 | 203 | 54/05 | 64/64 |
| | 2 | 901 | 359 | 357 | 215 | 39/84 | 60/22 |
| 7×7 | 0/1 | 584 | 477 | 215 | 130 | 81/67 | 60/46 |
| | 0/5 | 595 | 472 | 215 | 135 | 79/32 | 62/79 |
| | 1 | 514 | 448 | 218 | 144 | 72/96 | 66/05 |
| | 1/5 | 753 | 404 | 307 | 202 | 53/65 | 65/79 |
| | 2 | 901 | 357 | 360 | 217 | 39/62 | 60/27 |

### D. *Comparisons Using Histogram Differences*

After evaluations and examinations, the relationships mentioned have been compared in Table VI to improve the work, and the results have been shown below. Table IV and Table V show the different calculations of the histogram difference calculation, where $H_i$ represents the current frame histogram, $H_{i-1}$ denotes the previous frame histogram, *i* is the frame number and *n* is the last frame number, which are calculated for all frames. The following proposed Eqs. (18–21) introduce how to calculate histogram differences and their relationships in this research. The Fig. 7 exhibits results comparison using following relationships:

$$F2 = \sqrt{\sum_{i=1}^{n} (H_i - H_{i-1}) \times (H_i - H_{i-1})} \tag{18}$$

$$F2 = \sqrt{\sum_{i=1}^{n} (H_i - H_{i-1}) / (H_i - H_{i-1})} \tag{19}$$

$$F3 = \sum_{i=1}^{n} \sqrt{(H_i - H_{i-1}) \times (H_i - H_{i-1})} \tag{20}$$

$$F4 = \sum_{i=1}^{n} \sqrt{(H_i - H_{i-1}) / (H_i - H_{i-1})} \tag{21}$$

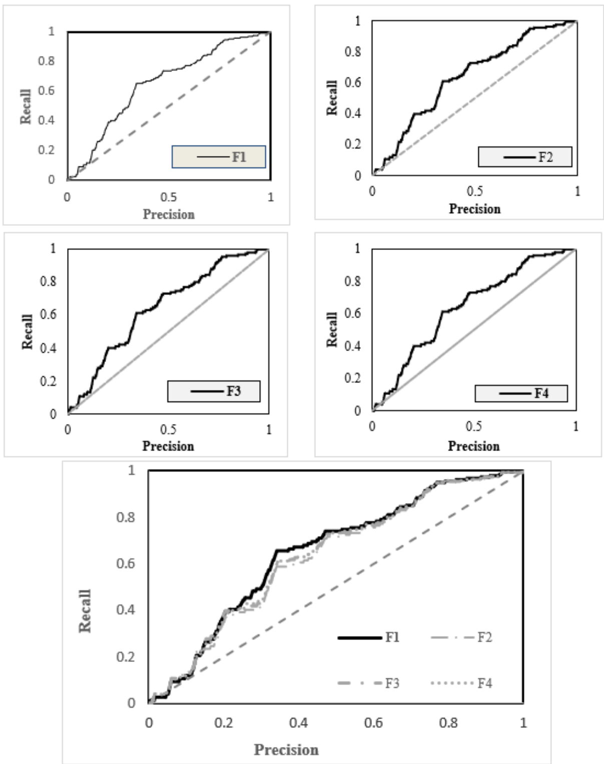


Figure 7. Results of the comparison of the mentioned relationships.

According to the results of the F1 relation, a better result has been provided because it gives us suitable hard cuts in the output, and we have fewer excessive cuts.

## V. Conclusion

Today, due to the huge volume of video data and the need to process, categorize and maintain them, using human resources to do such things is very time-consuming, costly, and even impossible. Shot boundary detection reduces unused data for our operations, and deleting this data speeds up processing. Moreover, this operation helps discover new information among large volumes of data. However, it should be noted that critical information can also be lost in some videos. This article has presented an approach to identify the boundary of shots in video sequences with the least errors. First, we removed the shots that did not include abrupt cuts with statistical and histogram difference methods. We have detected the faces at high accuracy in the shots with abrupt cut through the Viola-Joyce algorithm. As future work, we will explore the shot boundary in video sequences, which include gradual and abrupt changes in brightness in sequential shots, gradual fading, and rapid displacement of objects to increase the efficiency of shot boundary detection in video sequences.

## Conflict of Interest

The authors have no conflict of interest to declare.

## Author Contributions

A. R. Khan was involved in conceptualization, methodology, writeup and funding. M. Harouni contributed in methodology, writeup, analytics, and visualization. S. Sharifi was involved in methodology, experiments, results evaluation and drafting article. S. A. Bahaj worked on results validation, analysis, visualization and final review. T. Saba contributed to conceptualization, methodology, supervision, final review and editing. All authors had approved the final version.

## Funding

This research is technically supported by Artificial Intelligence & Data Analytics Lab (AIDA) CCIS Prince Sultan University, Riyadh, Saudi Arabia. The authors are also thankful for APC’s support of this publication.

## References

[1] M. Z. Khan, S. Jabeen, M. U. G. Khan, *et al.*, “A realistic image generation of face from text description using the fully trained generative adversarial networks,” *IEEE Access*, vol. 9, pp. 1250–1260, 2020.

[2] K. Meethongjan, M. Dzulkifli, A. Rehman, *et al.*, “An intelligent fused approach for face recognition,” *Journal of Intelligent Systems*, vol. 22, no. 2, pp. 197–212, 2013.

[3] S. Joudaki, D. B. Mohamad, T. Saba, *et al.*, “Vision-Based sign language classification: A directional review,” *IETE Technical Review*, vol. 31, no. 5, pp. 383–391, 2014.

[4] M. Harouni and H. Y. Baghmaleki, “Color image segmentation metrics,” *Encyclopedia of Image Processing*, vol. 95, pp. 10–21, 2018.

[5] T. Saba, A. Rehman, A. Altameem, *et al.*, “Annotated comparisons of proposed preprocessing techniques for script recognition,” *Neural Computing and Applications*, vol. 25, no. 6, pp. 1337–1347, 2014.

[6] A. M. Ahmad, S. Sulong, A. Rehman, *et al.*, “Data hiding based on improved exploiting modification direction method and Huffman coding,” *Journal of Intelligent Systems*, vol. 23, no. 4, pp. 451–459, 2014.

[7] M. A. Khan, T. Akram, M. Sharif, *et al.*, “An automated system for cucumber leaf diseased spot detection and classification using improved saliency method and deep features selection,” *Multimedia Tools and Applications*, vol. 79, no. 25, pp. 18627–18656, 2020.

[8] K. Neamah, D. Mohamad, T. Saba, *et al.*, “Discriminative features mining for offline handwritten signature verification,” *3D Research*, vol. 5, no. 1, pp. 1–6, 2014.

[9] A. Rehman, M. Harouni, N. H. S. Karchegani, *et al.*, “Identity verification using palm print microscopic images based on median robust extended local binary pattern features and k-nearest

neighbor classifier," *Microscopy Research and Technique*, vol. 85, pp. 1224–1237, 2022.
[10] G. Lenz, S. H. Ieng, and R. Benosman, "Event-based face detection and tracking using the dynamics of eye blinks," *Frontiers in Neuroscience*, vol. 587, 2020.
[11] M. S. M. Rahim, A. Rehman, F. Kurniawan, *et al.*, "Ear biometrics for human classification based on region features mining," *Biomedical Research*, vol. 28, no. 10, pp. 4660–4664, 2017.
[12] F. Afza, M. A. Khan, M. Sharif, *et al.*, "A framework of human action recognition using length control features fusion and weighted entropy-variances based feature selection," *Image and Vision Computing*, vol. 106, 104090, 2021.
[13] A. R. Khan, F. Doosti, M. Karimi, *et al.*, "Authentication through gender classification from iris images using support vector machine," *Microscopy Research and Technique*, vol. 84, pp. 2666–2676, 2021.
[14] R. Abinaya, L. P. Maguluri, S. Narayana, *et al.*, "A novel biometric approach for facial image recognition using deep learning techniques," *International Journal of Advanced Research in Engineering and Technology*, vol. 11, no. 9, 2020.
[15] R. Mishra, S. Singhai, and M. Sharma, "Comparative study of block matching algorithm and dual tree complex wavelet transform for shot detection in videos," in *Proc. International Conference on Electronic Systems, Signal Processing and Computing Technologies*, 2014, pp. 450–455.
[16] X. Ben, Y. Ren, J. Zhang, *et al.*, "Video-based facial micro-expression analysis: A survey of datasets, features and algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.
[17] W. Liu, "Sports games video segmentation based on adaptive," *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 9, no. 2, pp. 25–34, 2016.
[18] M. M. Dashti and M. Harouni, "Smile and laugh expressions detection based on local minimum key points," Arxiv Preprint Arxiv:2101.01874, 2021.
[19] R. Mishra, "Video shot boundary detection using hybrid dual tree complex wavelet transform with Walsh Hadamard transform," *Multimedia Tools and Applications*, vol. 80, pp. 28109–28135, 2021.
[20] G. Pal, D. Rudrapaul, S. Acharjee, *et al.*, "Video shot boundary detection: A review," in *Emerging ICT for Bridging the Future-Proceedings of the 49th Annual Convention of the Computer Society of India CSI*, Springer, 2015, vol. 2, pp. 119–127.
[21] A. Singh, D. M. Thounaojam, and S. Chakraborty, "A novel automatic shot boundary detection algorithm: Robust to illumination and motion effect," *Signal, Image and Video Processing*, vol. 14, pp. 645–653, 2020.
[22] C. Gupta, N. K. Gondhi, and P. K. Lehana, "Gray Level Co-occurrence Matrix (GLCM) parameters analysis for pyoderma image variants," *Journal of Computational and Theoretical Nanoscience*, vol. 17, no. 1, pp. 353–358, 2020.
[23] G. Zajic, A. Gavrovska, I. Reljin, *et al.*, "A video hard cut detection using multifractal features," *Multimedia Tools and Applications*, vol. 78, pp. 6233–6252, 2019.
[24] P. Gunawardena, O. Amila, H. Sudarshana, *et al.*, "Real-Time automated video highlight generation with dual-stream hierarchical growing self-organizing maps," *Journal of Real-Time Image Processing*, vol. 18, pp. 1457–1475, 2021.
[25] H. Bhaumik, S. Bhattacharyya, and S. Chakraborty, "A vague set approach for identifying shot transition in videos using multiple feature amalgamation," *Applied Soft Computing*, vol. 75, pp. 633–651, 2019.
[26] S. Chakraborty, A. Singh, and D. M. Thounaojam, "A novel bifold-stage shot boundary detection algorithm: Invariant to motion and illumination," *The Visual Computer*, vol. 38, pp. 445–456, 2022.
[27] J. M. Al-Tuwaijari and S. A. Shaker, "Face detection system based viola-jones algorithm," in *Proc. 6th International Engineering Conference "Sustainable Technology and Development"*, February 2020, pp. 211–215.
[28] M. Sharif, F. Naz, M. Yasmin, *et al.*, "Face recognition: A survey," J*ournal of Engineering Science & Technology Review*, vol. 10, no. 2, pp. 166–177, 2017.
[29] M. Elarbi-Boudihir, A. Rehman, and T. Saba, "Video motion perception using optimized Gabor filter," *International Journal of Physical Sciences*, vol. 6, no. 12, pp. 2799–2806, 2011.
[30] T. Saba, A. Rehman, and G. Sulong, "Improved statistical features for cursive character recognition," *International Journal of Innovative Computing, Information and Control*, vol. 7, no. 9, pp. 5211–5224, 2011.
[31] A. Nodehi, G. Sulong, M. Al-Rodhaan, *et al.*, "Intelligent fuzzy approach for fast fractal image compression," *EURASIP Journal on Advances in Signal Processing*, vol. 2014, no. 1, pp. 1–9, 2014.
[32] A. E. Rad, M. S. M. Rahim, A. Rehman, *et al.*, "Digital dental X-ray database for caries screening," *3D Research*, vol. 7, no. 2, pp. 1–5, 2016.
[33] S. Saleem, M. Khan, U. Ghani, *et al.*, "Efficient facial recognition authentication using edge and density variant sketch generator," *CMC-Computers Materials & Continua*, vol. 70, no. 1, pp. 505–521, 2022.
[34] F. Becattini, F. Palai, and A. D. Bimbo, "Understanding human reactions looking at facial micro-expressions with an event camera," *IEEE Transactions on Industrial Informatics*, 2022.



**Amjad Rehman Khan** is a senior researcher in the AIDA Lab CCIS Prince Sultan University Riyadh Saudi Arabia. He received his PhD & postdoc from Faculty of Computing, Universiti Teknologi Malaysia in information security with honor in 2010 and 2011 respectively. He received rector award for 2010 best student in the university. His interests are in information security, health informatics, pattern recognition.

**Majid Harouni** is assistant prof. in Department of Computer Engineering, Dolatabad Branch, Islamic Azad University, Isfahan, Iran. He earned PhD from Faculty of Computing, Universiti Teknologi Malaysia. His interests are in image processing & pattern recognition.

**Sepideh Sharifi** is also associated with Department of Computer Engineering, Dolatabad Branch, Islamic Azad University, Isfahan, Iran. Her interests are in image processing & pattern recognition.

**Saeed Ali Bahaj** is associate professor at MIS Department College of Business Administration, Prince Sattam bin Abdul-Aziz. He earned doctoral at Pune University India in 2006. His main research interests include information management, forecasting, information engineering and information security.

**Tanzila Saba** earned the PhD in document information security and management from Faculty of Computing, Universiti Teknologi Malaysia, Malaysia in 2012. She won best student award in the Faculty of Computing UTM for 2012. Currently, she is research Prof. in CCIS Prince Sultan University Riyadh Saudi Arabia. Her primary research focus in the recent years includes security, bioinformatics, classification.