Correlation Coefficients and Adaptive Threshold-Based Dissolve Detection in High-Quality Videos

Original Scientific Paper

[1]**Kamal S. Chandwani**
Sarvepalli Radhakrishnan University
Bhopal, Madhya Pradesh, India
chandwani1@rediffmail.com

[2]**Varsha Namdeo**
Department of Computer Science & Engineering
Sarvepalli Radhakrishnan University
Bhopal, Madhya Pradesh, India
varsha_namdeo@yahoo.com

[3]**Poonam T. Agarkar**
Department of Electronics & Telecommunication Engineering
Yeshwantrao Chavan College of Engineering
Nagpur, Maharashtra, India
poonamagarkar71@gmail.com

[4]**Sanjay M. Malode**
Department of Computer Science & Engineering
K. D. K. College of Engineering
Nagpur, Maharashtra, India
malodesanjay14@gmail.com

[5]**Prashant R. Patil**
Department of Management Studies
Smt. Radhikatai Pandav College of Engineering
Nagpur, Maharashtra, India
patilnagpur@gmail.com

[6]**Narendra P. Giradkar**
Department of Electronics & Telecommunication Engineering
Smt. Radhikatai Pandav College of Engineering
Nagpur, Maharashtra, India
giradkarnaresh@gmail.com

[7]**Pratik R. Hajare**
Mansarovar Global University
Bhopal, M.P., India
pratikhajare8@gmail.com

Abstract *– Rapid enhancements in Multimedia tools and features day per day have made entertainment amazing and the quality visual effects have attracted every individual to watch these days' videos. The fast-changing scenes, light effects, and undistinguishable blending of diverse frames have created challenges for researchers in detecting gradual transitions. The proposed work concentrates to detect gradual transitions in videos using correlation coefficients obtained using color histograms and an adaptive thresholding mechanism. Other gradual transitions including fade out, fade in, and cuts are eliminated successfully, and dissolves are then detected from the acquired video frames. The characteristics of the normalized correlation coefficient are studied carefully and dissolve are extracted simply with low computational and time complexity. The confusion between fade in/out and dissolves is discriminated against using the adaptive threshold and the absence of spikes is not part of the case of dissolves. The experimental results obtained over 14 videos involving lightning effects and rapid object motions from Indian film songs accurately detected 22 out of 25 gradual transitions while falsely detecting one transition. The performance of the proposed scheme over four benchmark videos of the TRECVID 2001 dataset obtained 91.6, 94.33, and 92.03 values for precision, recall, and F-measure respectively.*

Keywords*: Multimedia tools, gradual transitions, correlation coefficients, color histograms, adaptive thresholding, fade out, fade in, cuts, and dissolve*

1. INTRODUCTION

The challenges in video summarization and then retrieval had become the foremost need owing to rapid development in Multimedia technology. The development introduced extremely high-quality techniques in video processing and certainly made it difficult to extract meaningful segments from a video called shots. The greatest challenges introduced include uneven illuminations, camera movements, object rotations, low-contrast backgrounds, low-frequency transitions, etc. The most common approach used in video summarization involves pre-processing the video frames, extracting and selecting optimum features, finding correlation coefficients between successive or neighboring frames, detecting transitions, identifying key frames, detecting and tracking objects, classification, etc. Shots carry vital information regarding specific scenes and can be differentiated by detecting abrupt transitions and gradual transitions. Abrupt transitions are even perceptual and can be easily detected using simple algorithms when the correlation between successive frames is evaluated. Gradual transitions are hard to isolate due to several effects introduced such as dissolves, fade in and out, and wipes. Unlike abrupt transitions, the correlation between successive frames in the case of gradual transitions is very low that is the neighboring frames show high similarity. In detecting such gradual transitions, the transition effects have to be compensated and movements pertaining to objects and camera had to be cautiously handled. Various approaches as studied from the literature used key point detection in frames before finding the similarity between two successive frames. Patch matching with neighboring patches is another sort of technique to perfectly locate the moving patch or region in the frames but requires high computational and time complexity.

Recently, researcher have focused their research on dissolve detection due to unacceptably high false hit rate. Most of them are based on histograms, pixels, edges, etc. Histograms offer an advantage since they are invariant to local motions or small global motions. The authors in [1] used threshold to determine transitions when it was applied to the sum of the absolute difference between two successive frames. The transition was declared when the sum was greater than the threshold. A similar approach was used by [2-4]. The work suggested by [5] used an adaptive threshold instead of a fixed threshold based on a sliding window. Region-based matching was proposed in [6] to determine the best match between two neighboring frames. The frames were divided into 12 patches or regions to ensure the extraction of motion vectors and the weighted sum of sorted pixel difference was considered as the correlation between two frames. A twin comparison method [4] was introduced to compare two histograms based on the histogram difference metric. A block-based histogram over RGB color frames was used in [7]. The method in [8] used histograms with 64 bins and block-based histograms in two stages. In the first stage, global histograms were subtracted and their absolute value was compared with a threshold while in the subsequent stage, the histogram difference metric was compared. The number of region differences exceeding the difference threshold when exceeded the count threshold, a transition was observed [9-10]. An inter-frame difference histogram based on the Poisson model to detect abrupt and gradual transitions was proposed in [11]. The work proposed in [12] introduced scale-invariant feature transform to RGB color space while the method in [13] compared different Shot boundary detection schemes using color histograms as features. HSV histogram with level 5 difference was used to detect a gradual transition in [14].

This paper is organized with the following headings: Related work covers state of art work contributed by different researchers: Methods and Materials deals with the proposed work and covers algorithm and complete description of the system: Results and Discussion includes experimental results obtained over two different dataset videos under test and performance parameters: and Conclusion focuses on the summary of the work, loopholes and future scope for the work.

The paper is organized in the following sections. Section 2 summarizes state of art work by other researchers. Section 3 deals with the dataset and the proposed methodology to detect dissolve transitions. Section 4 includes results and discussion while the last section 5 concludes the work.

2. RELATED WORK

Frame features were extracted using principal component analysis to uplift prominent features and then subjected to distance calculating algorithm [15]. The detection accuracy was then improved by reviewing false detection boundaries to convolutional neural networks. The work by [16] suggested improving false detections of transitions and extracting multiple invariant features such as scale-invariant feature transform, color layout descriptor, and edge change ratio. These features compensate for the effect of variations in illumination, motion, scaling, and rotation. The feature set was classified using the SVM classifier to obtain an F1-score of 97% over the TRECVid 2001 dataset. The frames were segmented in [17] into primary and candidate segments to analyze the transition behavior in frames. Segmentation was obtained using the color feature and local adaptive threshold of each part or segment. Speeded Up robust features extracted from boundary segments were utilized to fine-tune the transitions from the candidate segments over TRECVid 2001 dataset to achieve a 90.8% F1-score in case of gradual transitions. A two-stage shot detection technique was presented in [18] employing a fusion of color histogram and deep features for distinguishing hard cuts and C3D-based deep analysis to locate gradual cuts. Abrupt shots are used to partition the video into segments and 3D-convolutional neural networks are used to classify the clips into specific gradual transitions. Merging techniques were used to know the positions of the gradual transitions.

Work introduced in [19] used the three-stage process to detect candidate boundaries, and instant and gradual transitions through deep CNN. They improved the performance by filtering out many non-boundaries and introduced a scheme to locate the start, mid and end points of the gradual transitions. Authors in [20] introduced a two-phased approach to detect candidate dissolves and filter candidates based on a threshold. The first stage involves identifying parabolic patterns in the mean fuzzy entropy of the frames and the second stage uses an ensemble of four parameters for filtering, respectively. The work suggested in [21] detected abrupt shots using Binarized edge information through a linear binary pattern and estimated the Euclidean distance of the histogram features and then applied an adaptive threshold. For key frame extraction, a Sobel operator was used to get the magnitude gradient of frames under a segmented shot and transformed into z-scores. The frame possessing the highest value corresponding to the coefficient of variation was selected as the key frame for the shot.

2.1. OUR CONTRIBUTION

We propose a simple and efficient method to detect dissolve transitions in high-quality videos using correlation coefficients obtained from color histograms. The approach includes:

1. Calculation of Adaptive threshold based on correlation coefficients found over color histograms for isolating other transitions (fade in and fade out) from dissolve transitions.
2. Automatic detection of dissolved transition frames based on width and subsequently the peaks inside the transition.

3. PROPOSED WORK

The proposed work is concentrated on detecting dissolve transitions in videos based on calculating the consecutive frame differences and then thresholding the histogram. The histogram is obtained from the correlated values between two consecutive frames. Histograms for all three color components of two consecutive frames are considered for evaluating the correlated values. It is obtained by subtracting the histograms of individual color components, followed by taking the absolute values, and finally finding the sum. The average value of all three color components across the video clip is then averaged and normalized using the maximum value, and a histogram corresponding to (N-1) correlated values is obtained. The maximum values in the (N-1) correlated values are used as a threshold for the (N-1) values. Finally, the threshold correlated values are normalized again concerning the maximum value. The flowchart of the proposed algorithm for detecting the dissolve transitions is shown in Fig. 1 below. N represents the total number of frames in the clip or video.

Considering the $frame_m$ to represent one of the two consecutive frames, the histogram h_c for any of the color components is given by expression (1), where H represents the histogram.

$$h_c = H\,(frame_m)_{\,frame\,\epsilon\,(R,\,G,\,B)} \tag{1}$$

The correlation between two consecutive frames is thus obtained by the following expression (2), where (i) and (i+1) are any two consecutive frames, and m represents any of the channels including the R, G, and B.

$$C_{m\ \epsilon\,(R,\,G,\,B)} = sum\ (abs\ (h_{im} - h_{(i+1)m})) \tag{2}$$

The correlated values are stored in an array D which can be represented by,

$$D = \cup_{m=1}^{n=N-1}[C_R; C_B; C_G] \tag{3}$$

The correlation coefficients are obtained by taking the average value of all three components and expressed using the following equation (4),

$$M = \frac{1}{3}\,\sum_{i=1}^{N-1} D_{R,G,B} \tag{4}$$

The normalizing coefficient M_{max} is obtained by finding the maximum value from M. Equation (5) finds the maximum value.

$$M_{max} = Max\ (M) \tag{5}$$

The correlation coefficients are normalized using the expression (6),

$$F = \frac{M}{M_x} \tag{6}$$

Now we find the histogram of the normalized correlation coefficients using expression (7).

$$G = H\ (F) \tag{7}$$

The adaptive threshold is then calculated using the maximum value in the histogram which is given by the following equation (8).

$$T = Max\ (G) \tag{8}$$

With this adaptive threshold T, the values in F are subjected to the threshold using the expression (9).

$$S = \begin{cases} 1 & if\ F > T \\ 0 & Otherwise \end{cases} \tag{9}$$

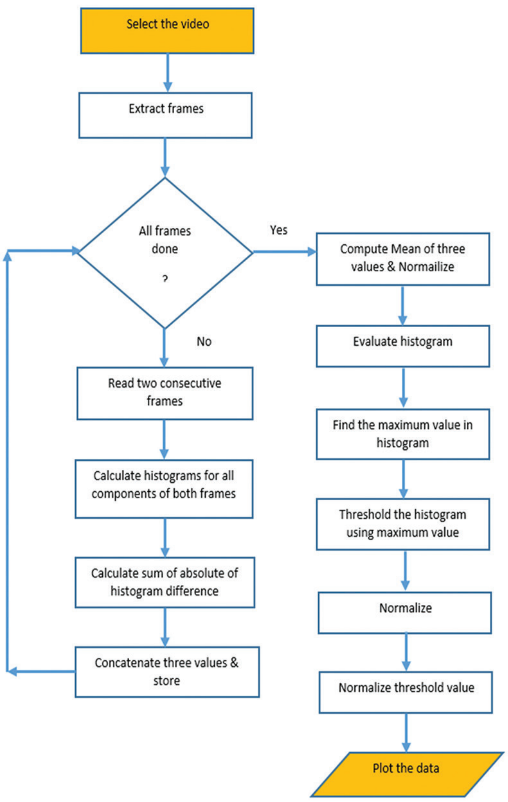


Fig. 1. Flowchart for the proposed algorithm to detect the Dissolve transitions

The original values remaining after thresholding are obtained by multiplying the result obtained using equation (9) by F again, which can be represented by equation (10).

$$S = S\ x\ F \tag{10}$$

Finally, the threshold T is made consistent with the values in S by dividing the threshold T by value 256.

$$T = T/256 \tag{11}$$

Histogram of normalized values is obtained further and the maximum value in the histogram is considered as a threshold. The threshold value is used to threshold the values of the histogram and multiplied by values in the histogram again. The threshold value is normalized by dividing it by value 256 to make it consistent with the histogram obtained in the previous stage.

The following Fig. 2 depicts the plot for one of video clips 4 from the Ashiqui2 song.

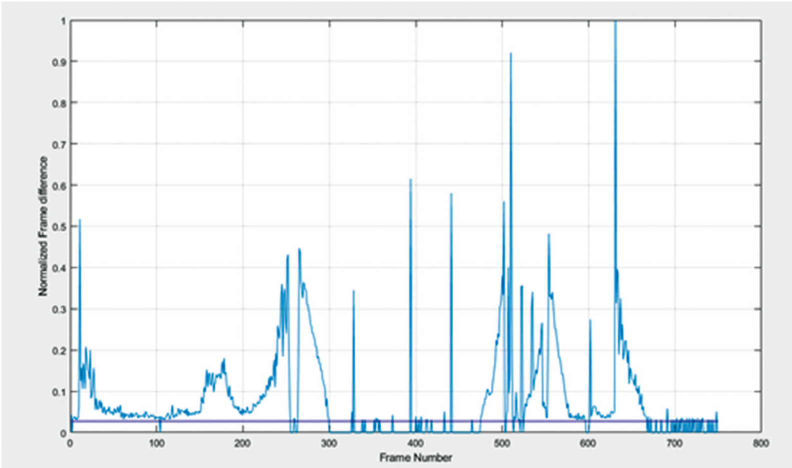


Fig. 2. Final plot obtained by our proposed algorithm on video clip 4 with a horizontal blue line as adaptive threshold

The following algorithm explains the steps to find the correlation coefficients and the adaptive threshold to detect gradual transitions in high-quality videos.

Algorithm 1 – Dissolve Detection

Input: Video Clip V

Output: Frames containing Dissolves F_n

Find the number of frames in the video clip V; N

For m =1 to N-1; N-Number of frames in the video clip

  Read two successive frames

  Evaluate the histogram for each color component; h

  Calculate the sum of absolute values for each component; C

  Store the sum of color component sum; D

End loop; m

Compute the Mean of each color component M and Normalize

Evaluate the histogram from the normalized mean value; G

Compute the Adaptive Threshold; T

Find the Occurrence of gradual transitions

Differentiate Fades and Dissolve Transitions based on Peaks

Find the start and end frames involving dissolve transitions; F_n

After eliminating the fade transitions, the transitions above the threshold line are found using the threshold. The starting frame is noted concerning the transition from low to high transition concerning the threshold line. The transition should remain high for at least 6 frames for a true dissolve. The cut-off value of 6 frames is found from experimental analysis and found to agree with more than 98% of cases. The other 2% of cases include fast camera motion accompanied by low illuminations and are confused with dissolves. The ending frame involves a high-to-low transition concerning the threshold. Similarly, other gradual transitions are found.

3.1. DATASET

We examined two video songs from Hindi movies *Ashiqui2* and *Once Upon a Time in Mumbai* and the other four videos from the benchmark dataset TRECVID 2001 for comparison. The song videos were cut into smaller videos consisting of 500 to 800 frames. The reason behind this is to reduce the computations and time complexity. The videos are so selected to test the effectiveness of our proposed algorithm since the songs are subjected to many objects plus camera motions and varying illuminations and include many dissolve, fade in, fade out, and cuts transitions. We also separated video clips involving wipe transitions but not covered their detection in this work. Fig. 3 shows a short dissolve transition from the song *Ashiqui2*. The number of frames is displayed at a slower rate reduced by 5 but the actual number of frames containing dissolve is 13, which starts with the 696th frame and ends at the 715th frame. All the videos have an original frame rate of 25 frames per second.

The video clips used in this work and the manual results obtained for dissolve transitions are given in Table 1 in the result and discussion section. For simplicity, we have not mentioned other transitions as those are not part of this work such as fade, fade-outs, and cuts. We have dealt with fade-in and fade-out in our earlier work [22]. We obtained 100% accuracy in detecting cuts, fade-in, and fade-out transitions.

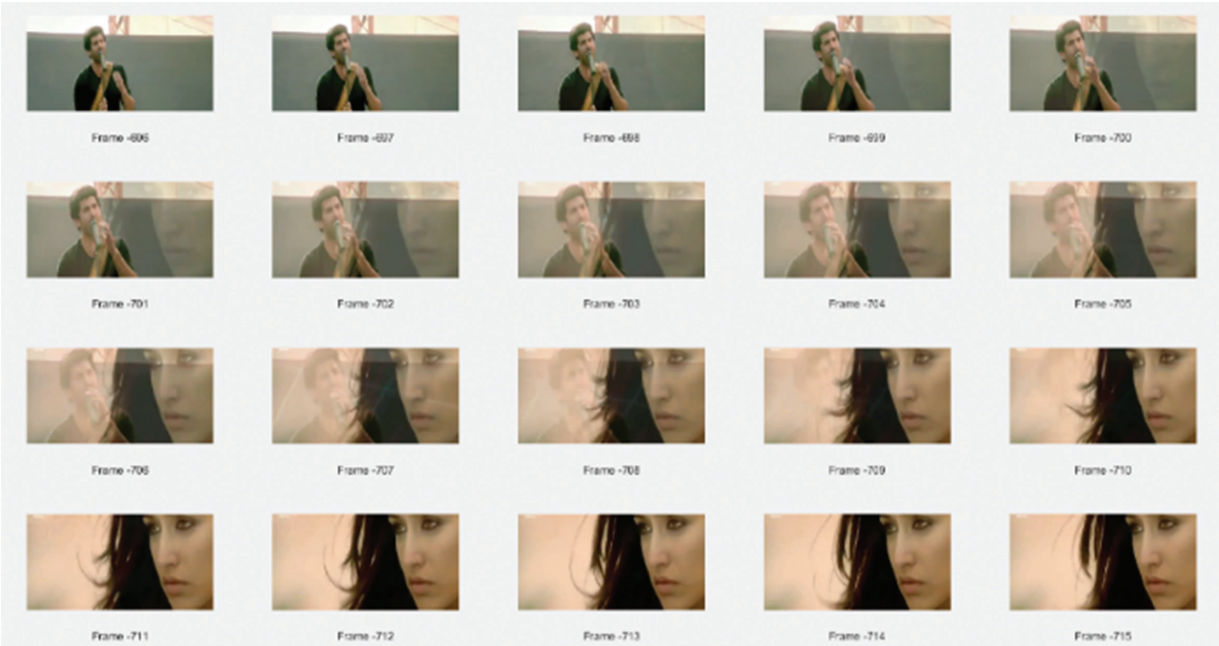


Fig. 3. A short dissolve (13 frames length) from the HINDI movie *ASHIQUI2* from frames 696 to 715

4. RESULTS AND DISCUSSION

Experimental evaluations carried out on 14 different clips containing rapid object/camera motions and uneven illumination showed that our technique was able to detect the gradual transitions with high accuracy except in some cases. The nature of variation in normalized correlation coefficients using the histogram technique as discussed earlier is presented in Fig. 4. The correlation coefficients touch the adaptive threshold line during the start and end of the gradual transition and remain above the threshold during the dissolve transition. The same effect is observed in the case of fade-in and fade-out transitions but there is an abrupt transition during the start and end frame which can be seen in figure 4. The effect of object motion and uneven illuminations are compensated using this technique. Slight spikes resulting from the object motion in dissolve transitions are unable to reach the threshold and thus can be detected as dissolves.

As seen from Figure 4, fade out followed by fade in as acquired by the correlated coefficients have different characteristics as compared to dissolve transitions. Dissolve transitions never touches the adaptive threshold line (Red) during the complete transition. Cuts are one-frame spikes as indicated by the green arrow. The small spikes occurring between the start and end of any of the transitions are effects of object motion or due to differences of illuminations between successive frames. The following figures from 5 to 11 represent the nature of correlation coefficient plots obtained for other videos.

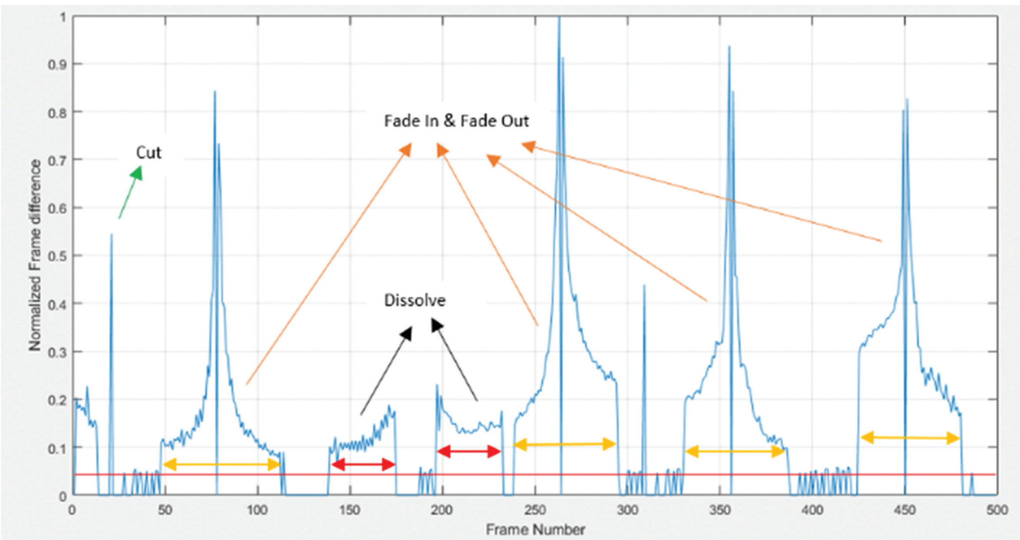


Fig. 4. Final plot obtained by our proposed algorithm on video clip 13. The correlation coefficient characteristics for cut (green), dissolve (black), and fade in and fade out (orange).

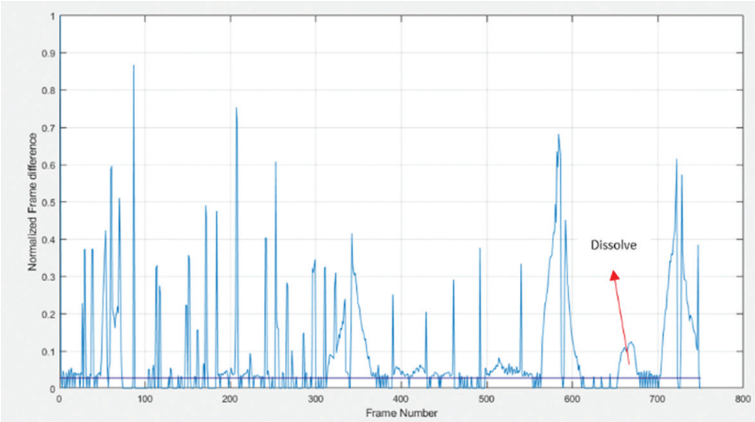


Fig. 5. Video clip 1 and the detected dissolve

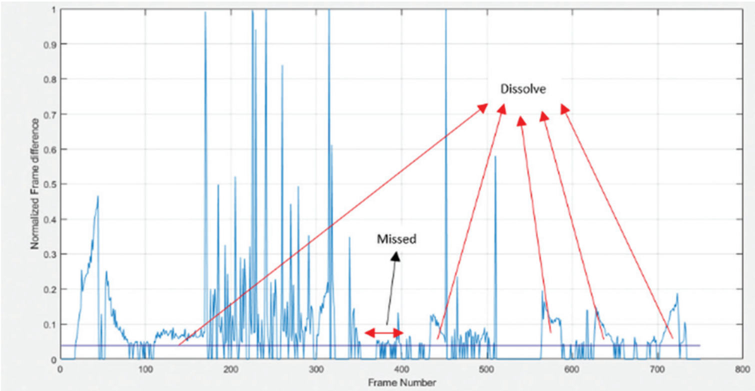


Fig. 6. Video clip 2 with Missed and detected dissolves

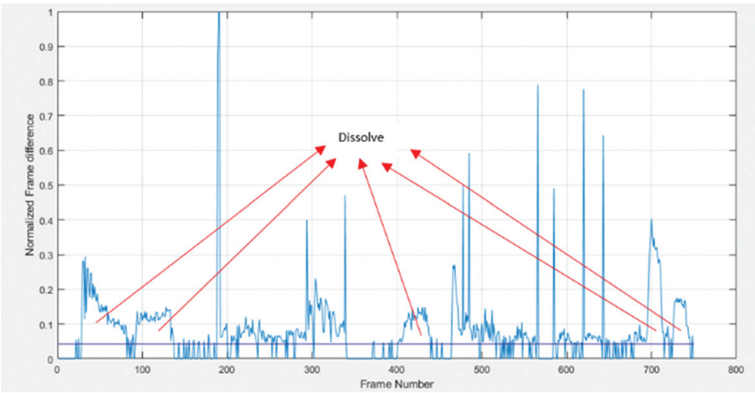


Fig. 7. Video clip 3 with detected dissolves

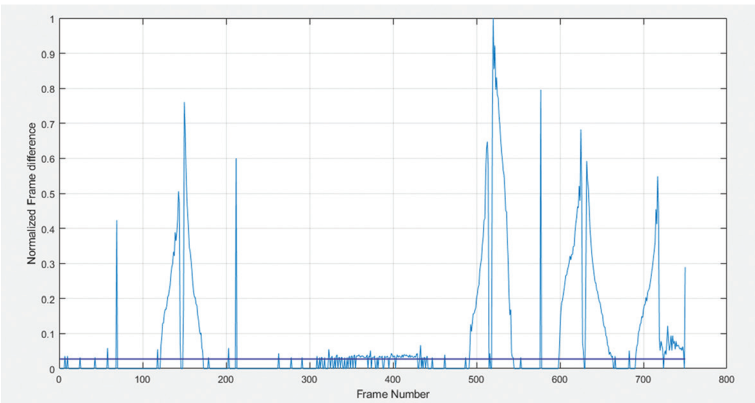


Fig. 8. Video clip 5 with no dissolve transitions

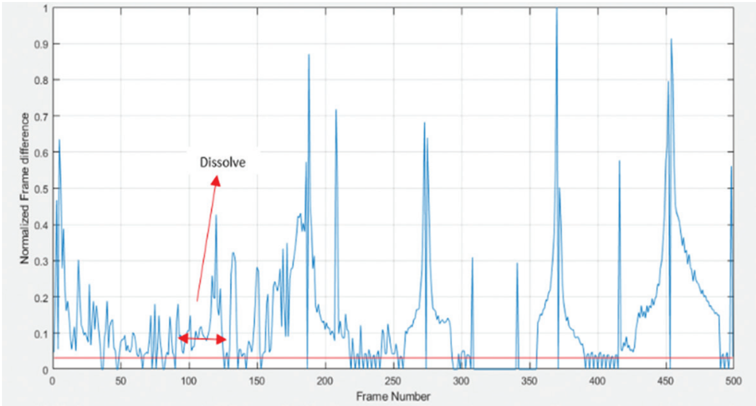


Fig. 9. Video clip 12 with no dissolve transitions

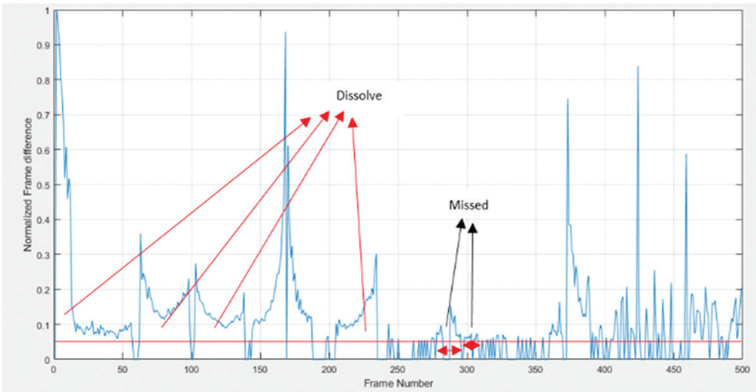


Fig. 10. Video clip 15 with Missed and detected dissolves

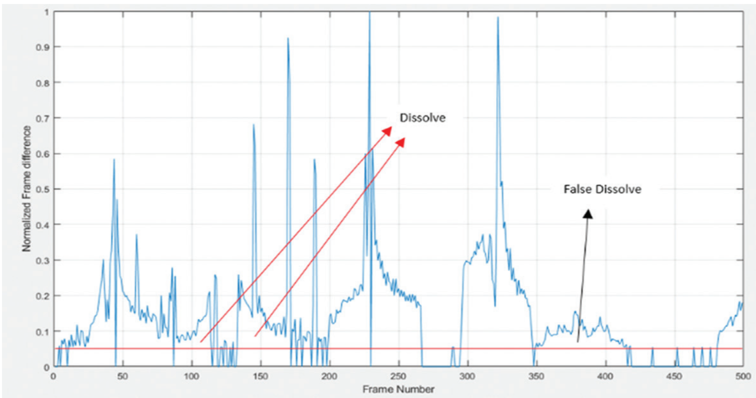
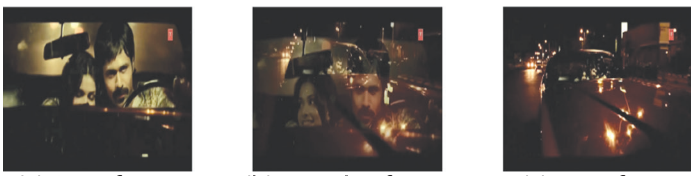


Fig. 11. Video clip 16 with False and True dissolves

Video Clips 6, 7, 8, 9, 10, 11, 13, 14, and 17 are not shown since they have no dissolve transitions and video Clip 18 has two dissolve transitions but are accurately detected by our proposed system. Video clip 2 (Missed dissolve from frame 380 to 400) shows missed dissolved transitions due to the reasons that there exist extreme lightning changes and rapid camera motion. The undetected dissolves in video clip 15 (244-280 & 290-325) were due to densely populated objects and very slow transitions of similar types of frames at the beginning and end of the transition. The similarity is due to the old frame being closely associated with the new frame as shown in Fig. 12.



(a) Start frame (b) Dissolve frame (c) Next frame

Fig. 12. Dissolve scene from video clip 15, where the start frame and the next frame are closely associated concerning illumination and background (dark)

The false dissolve in video clip 16 is due to the motion of a large object covering almost 80% of the significant frame. Advancements in multimedia technology no matter had improved the perceptual quality of videos and made the entertainment eye-catching but at the same time increased the challenge of detecting gradual transitions and was almost impossible in some cases where the blending is done within a few frames (less than 5-6 frames). Table 1 shows the results obtained using our technique.

Table 1. Comparison between dissolve detected with the proposed technique and the ground truth

Sr. No.	Video Song	Video Clip	Actual Dissolve transition (Ground Truth)	Dissolve transition detected	Discrepancies
1	Aashiqui 2	Video Clip 1	655-690	✓	
2		Video Clip 2	110-130 380-400 431-450 561-585 630-650 700-721	✓	Missed – (380-400)
3		Video Clip 3	30-80 100-135 413-440 696-711 728-750	✓	
4		Video Clips 4,5 and 6	None	✓	
5	Once upon a time in Mumbai	Video Clip 11, 14 and 17	None	✓	
6		Video Clip 12	99-120	✓	
7		Video Clip 13	140-175 197-235	✓	
8		Video Clip 15	25-55 64-100 105-140 204-235 244-280 290-325	✓	Missed – (244-280, 290-325)
9		Video Clip 16	81-110 135-120	✓	False dissolve – (345-420)
10		Video Clip 18	1-20 618-645		

Assuming TP for True positive samples, FP for false positive samples, and FN for false negative samples.

$$\text{Then, } Precision\ P = \frac{TP}{(TP+FP)} \tag{12}$$

$$Recall\ R = \frac{TP}{(TP+FN)} \tag{13}$$

$$\text{And, } F\text{-}Measure\ F = \frac{2*P*R}{(P+R)} \tag{14}$$

Therefore, $TP = 22$, $FP = 1$, $FN = 3$ gives

$P = 22/24 = 0.9167$,

$R = 22/(22+3) = 0.88$ and

$F = (2x0.9167x0.88)/(0.9167 + 0.88) = 0.90$.

The performance of the proposed scheme is compared with other state of artwork techniques based on quantitative analysis concerning the standard TECHVID 2001 dataset and includes Eigenvalue decomposition and Gaussian transition detection method [23], Walsh-Hadamard transforms kernel-based method [24], temporal segmentation method [25], Multimodal visual feature method [26], 3D convolutional network method [27], visual color information [28], Adaptive thresholds and gradual curve point [29] and SURF feature descriptor [30]. The work proposed in [26] obtained multimodal features using frame-based SURF features thus increasing complexity and ignoring the illumination changes. Work done in [28] however tracked the illumination variations in the L*a b color space but was unable to cope with tracking due to poor features on account of rotation and scaling. The adaptive threshold estimated in [30] is estimated over the full-length video which affects the generalization ability of the algorithm when videos involve rapid scene changes.

The work suggested in [29] relies on the average edge image. Deep fine edge detection operators increase the chances of misclassification and considering only hard edges increases the probability of transition elimination. The analysis in Table 2 shows that the proposed scheme is comparable concerning a precision score of 91.6 with the SURF feature-based approach in [30] while it outperforms in terms of Recall and F1-measure.

Table 2. A Quantitative Analysis of the Proposed Method on the TRECVID 2001 Dataset

Method	Video	Precision	Recall	F1-Score
Eigen Value Decomposition and Gaussian Transition Detection Method [23]	anni006	85.2	93.8	89
	anni009	82.1	84.5	81.5
	anni010	78.4	91.2	83.5
	nad58	92	93	91.4
	Average	**84.5**	**92.3**	**87**
Walsh–Hadamard Transform Kernel-Based Method [24]	anni006	90	87.1	88.5
	anni009	88.7	85.9	87.3
	anni010	84.6	80	82.2
	nad58	88.5	88.5	88.5
	Average	**88.0**	**85.4**	**86.6**
Temporal Segmentation Method [25]	anni006	92.2	84.3	91.1
	anni009	86	87.9	84.2
	anni010	84.2	88.5	85.0
	nad58	91.5	92.8	93.7
	Average	**86.7**	**89.1**	**89.0**
Multi-Modal Visual Features Method [26]	anni006	77.8	90.3	83.6
	anni009	89.6	67.2	76.8
	anni010	65.5	65.5	65.5
	nad58	80.4	82.2	81.3
	Average	**78.3**	**76.3**	**76.8**
3D Convolutional Network Method [27]	anni006	95.5	92.8	90.2
	anni009	80.6	94.3	88.2
	anni010	85.6	84.5	88.9
	nad58	90.6	90.2	92.1
	Average	**87.3**	**91.7**	**89.7**
Visual Colour Information Method [28]	anni006	83.3	88.2	85.7
	anni009	84.8	87.5	86.2
	anni010	81.7	89.1	85.2
	nad58	90.9	87	88.9
	Average	**85.2**	**88.0**	**86.5**
Adaptive Thresholds And Gradual Curve Point [29]	anni006	77.3	83.9	83.6
	anni009	90.2	85.5	87.5
	anni010	93.6	87.3	92.4
	nad58	94.1	90.2	92.5
	Average	**91.7**	**86.5**	**89.6**
SURF Feature Descriptor [30]	anni006	92.3	88.9	90.6
	anni009	92.7	91.1	91.9
	anni010	89.1	100	94.2
	nad58	92.7	90.5	91.6
	Average	**91.7**	**92.6**	**92.1**
Proposed Scheme	anni006	92.3	90.6	90.0
	anni009	91.6	100	90.4
	anni010	89.5	95.4	94.6
	nad58	93.1	91.3	93.1
	Average	**91.6**	**94.33**	**92.03**

The color depth-based approach using the correlation between successive frames can track illumination changes and camera motions. However, our approach failed when the gradual transition is fast and abrupt involving a transition of less than 6 frames. Also, dissolve transitions occurring in high proportionate dark backgrounds with low-intensity frames are missed since the frame lacks proper color features.

5. CONCLUSIONS

Out of 18 video clips, with 25 dissolve transitions, the novel approach presented in this work was able to detect dissolve transitions in 11 videos having 22 dissolves with 100% accuracy. The video clips 4, 5, 6, 11, 14, and 17 were accurately determined by our technique which contained no dissolve transitions. The reason for missed detection and false detection for videos 2, 15, and 16 had been quoted in the previous section. Thus our system for detecting gradual transitions in videos was successful in detecting 22 out of 25 transitions with an accuracy of approximately 92%. The precision, recall, and F-measure are 0.9167, 0.88, and 0.90 respectively. The detection can be improved if the effect of illumination changes is compensated using any of the state of art contrast correction methods and the effect of object motion is properly handled using the robust correlation between neighboring frames. The proposed work achieved remarkable results over the TRECVID 2001 dataset outperforming another state of the artwork. The precision, recall, and F1 measure were found to be 91.6, 94.33, and 92.03 respectively average over four videos from the dataset. Future work will be concentrated to improve the detection accuracy and eliminate false detection using extracting optimum features from the frames for finding correlations between neighboring frames and deducing the feature vector for optimum detection.

6. REFERENCES

[1] M. S. Lee, Y. M. Yang, S. W. Lee, "Automatic video parsing using shot boundary detection and camera operation analysis", Pattern Recognition, Vol. 34, 2001, pp. 711-719.

[2] T. Kikukawa, S. Kawaguchi, "Development of an automatic summary editing system for the audio-visual resources", The Transactions of the Institute of Electronics, Information and Communication Engineers, Vol. 75, 1992, pp. 204-212.

[3] A. Nagasaka, Y. Tanaka, "Automatic video indexing and full-video search for object appearances. In Visual Database Systems II", North-Holland Publishing Co., 1992, pp. 113-127.

[4] H. Zhang, A. Kankanhalli, S. W. Smoliar, "Automatic partitioning of full-motion video", Multimedia Systems, Vol. 1, 1993, pp. 10-28.

[5] B. L. Yeo, B. Liu, "Rapid Scene Analysis on Compressed Video", IEEE Transaction on Circuits and Systems, Video Technology, Vol. 5, 1995, pp. 533-544.

[6] B. Shahraray, "Scene change detection and content-based sampling of video sequences", Proceedings of the IS&T/SPIE's Symposium on Electronic Imaging: Science & Technology; International Society for Optics and Photonics, San Jose, CA, USA, 1995, pp. 2-13.

[7] D. Swanberg, C. F. Shu, J. R. Jain, "Knowledge-guided parsing in video databases", Proceedings of the IS&T/SPIE's Symposium on Electronic Imaging: Science and Technology; International Society for Optics and Photonics, San Jose, CA, USA, 1993, pp. 13-24.

[8] J. S. Boreczky, L. A. Rowe, "Comparison of video shot boundary detection techniques", Journal of Electronics and Imaging, Vol. 5, 1996, pp. 122-128.

[9] J. Yuan, H. Wang, L. Xiao, W. Zheng, J. Li, F. Lin, B. Zhang, "A formal study of shot boundary detection", IEEE Transaction on Circuits & Systems for Video Technology, Vol. 17, 2007, pp. 168-186.

[10] S. H. Abdulhussain, A. R. Ramli, M. I. Saripan, B. M. Mahmmod, S. A. R. Al-Haddad, W. A. Jassim, "Methods and challenges in Shot boundary detection: a review", Entropy, Vol. 20, No. 4, 2018, p. 214.

[11] Y. Huo, Y. Wang, H. Hu, "Effective algorithms for video shot and scene boundaries detection", Proceedings of the IEEE/ACIS 15th International Conference on Computer and Information Science, Okayama, Japan, 26-29 June 2016, pp. 1-6.

[12] Z. El Khattabi, Y. Tabii, A. Benkaddour, "Video shot boundary detection using the scale-invariant feature transform and RGB color channels", International Journal of Electrical and Computer Engineering, Vol. 7, 2017, pp. 2565-2573.

[13] U. Gargi, R. Kasturi, S. H. Strayer, "Performance characterization of video-shot-change detection methods", IEEE Transaction on Circuits & Systems for Video Technology, Vol. 10, No. 1, 2000, pp. 1-13.

[14] Z. Li, X. Liu, S. Zhang, "Shot boundary detection based on the multilevel difference of color histograms", Proceedings of the First International Conference on Multimedia and Image Processing, Bandar Seri Begawan, Brunei, June 2016, pp. 15-22.

[15] D. Chakraborty, W. Chiracharit, K. Chamnongthai, "Semantically Relevant Scene Detection Using Deep Learning", Proceedings of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, Tokyo, Japan, 14-17 December 2021, pp.1576-1579.

[16] T. J. Jose, S. Rajkumar, M. R. Ghalib, A. Shankar, P. Sharma, M. R. Khosravi, "Efficient Shot Boundary Detection with Multiple Visual Representations", Mobile Information Systems, Vol. 2022, 2022.

[17] M. R. Suguna, A. Kalaivani, S. Anusuya, "The Detection of Video Shot Transitions Based on Primary Segments Using the Adaptive Threshold of Colour-Based Histogram Differences and Candidate Segments Using the SURF Feature Descriptor", Symmetry, Vol. 14, No. 10, 2022, p. 2041.

[18] L. Wu, S. Zhang, M. Jian, Z. Lu, D. Wang, "Two-stage Shot Boundary Detection via Feature Fusion and Spatial-Temporal Convolutional Neural Networks", IEEE Access, Vol. 7, 2019, pp. 77268-77276.

[19] T. Wang, N. Feng, J. Yu, Y. He, Y. Hu, Y.-P. P. Chen, "Shot boundary detection through Multi-stage Deep Convolution Neural Network", Proceedings of the 27th International Conference, Multimedia and Multimodal Analytics in the Medical Domain and Pervasive Environments, Prague, Czech Republic, 22-24 June 2021.

[20] B. Hrishikesh, M. Chakraborty, S. Bhattacharya, S. Chakraborty, "Detection of Gradual Transition in Videos: Approaches and Applications", Intelligent Analysis of Multimedia Information, IGI Global, 2017, pp. 282-318.

[21] H. M. Nandini, H. K. Chethan, B. S. Rashmi, "Shot-based key-frame extraction using edge-LBP approach", Journal of King Saud University, Computer and Information Science, Vol. 34, 2022, pp. 4537-4545.

[22] K. Chandwani, V. Namdeo, N. Giradkar, P. Patil, "Multilevel wavelet-based features for detecting gradual transitions in High-Quality Videos", NeuroQuantology, Vol. 20, No. 13, 2022, pp. 2805-2812.

[23] A. Amiri, M. Fathy, "Video shot boundary detection using generalized eigenvalue decomposition and Gaussian transition detection", Computing and Informatics, Vol. 30, No. 3, 2012, pp. 595-619.

[24] G. G. L. Priya, S. Domnic, "Walsh-Hadamard Transform Kernel-Based Feature Vector for Shot Boundary Detection", IEEE Transaction on Image Processing, Vol. 23, 2014, pp. 5187-5197.

[25] E. Santos, A. C. Sousa, H. Pedrini, "Shot boundary detection for video temporal segmentation based on the Weber local descriptor", Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, Banff, AB, Canada, 5-8 October 2017, pp. 1310-1315.

[26] S. Tippaya, S. Sitjongsataporn, T. Tan, M. M. Khan, K. Chamnongthai, "Multi-Modal Visual Features-Based Video Shot Boundary Detection", IEEE Access, Vol. 5, 2017, pp. 12563-12575.

[27] T. Liu, Y. Lu, X. Lei, L. Zhang, H. Wang, W. Huang, Z. Wang, "Soccer video event detection using 3D convolutional networks and shot boundary detection via deep feature distance", Proceedings of the International Conference on Neural Information Processing, Guangzhou, China, 14-18 November 2017, pp. 440-449.

[28] S. Chakraborty, D. M. Thounaojam, N. Sinha, "A Shot boundary Detection Technique based on Visual Colour Information", Multimedia Tools & Applications, Vol. 80, 2021, pp. 4007-4022.

[29] N. Kumar, D. Raj, "Shot Boundary Detection Framework for Video Editing Via Adaptive Thresholds and Gradual Curve Point", Turkish Journal of Computer and Mathematics Education, 2021, Vol. 12, No. 11, pp. 3820-3828.

[30] M. Raja Suguna, A. Kalaivani, S. Anusuya, "The Detection of Video Shot Transitions Based on Primary Segments Using the Adaptive Threshold of Colour-Based Histogram Differences and Candidate Segments Using the SURF Feature Descriptor", Symmetry, Vol. 14, 2022, p. 2041.