A Novel Hybrid Keypoint Detection Algorithm for Gradual Shot Boundary Detection

Shunmugam Karpagavalli, Vanniappan Balamurugan and Sankaran Ramesh Kumar

February 17, 2020

A Novel Hybrid Keypoint Detection Algorithm for Gradual Shot Boundary Detection

Shunmugam Karpagavalli
Dept. of Computer Science and Engg.
Manonmaniam Sundaranar University
Tirunelveli, Tamilnadu, India, 627 012
srikarramash@gmail.com

Vanniappan Balamurugan
Dept. of Computer Science and Engg.
Manonmaniam Sundaranar University
Tirunelveli, Tamilnadu, India, 627 012
bala_vm@msuniv.ac.in

Sankaran Ramesh Kumar
Dept. of Computer Science and Engg.
Manonmaniam Sundaranar University
Tirunelveli, Tamilnadu, India, 627 012
mailtodr.rameshkumar@gmail.com

***Abstract* — In the recent period, video summarization plays a vital role in the video retrieval process especially in the video surveillance system. The summarization process comprises of several activities such as preprocessing, features extraction, frame similarity analysis, shot boundary detection (SBD), etc. This paper presents a novel hybrid keypoint detection algorithm for the gradual shot boundary detection. The proposed technique combines the Hessian matrix of point of interest and the minimum Eigen values of the region of interest. The proposed technique has been tested on several video data sets and the results shows better performance in terms of accuracy when compared to the previous methods such as MEV, CD, FAST, SURF, MSER and BRISK techniques.**

Keywords— Video summarization, shot boundary detection, Gradual Transition, Eigen values

I. Introduction

Video summarization is a process of comprehending the content of the video files by selecting the key frames which are decided by the features of the frame. The summarized frame sequences present a meaningful abstract view of the entire video within a short duration. Video summarization process involves several sub processes viz. preprocessing, features extraction, frame similarity analysis, shot boundary detection, key frame identification, object detection, tracking, classification, etc. Shot boundary detection plays a vital role in extracting and organizing the key frames in the summarization process, as the video sequence consists of several shots [1]. In general, shots can be categorized as Abrupt Transition (AT) and Gradual Transition (GT) [2]. Abrupt Transition can be easily detected by comparing the content variations between the two consecutive frames, whereas, the GT needs several steps to identify the boundary as they have several transition effects such as dissolves, fades, and wipes [3], [4]. Several researchers have focused on AT by applying two-step approach viz. frame segmentation [5] and correlation analysis [6]. Unlike AT, the GT between two consecutive shots usually lasts for a while. Also, in case of GT, the similarity distance between any two consecutive frames is negligible. Therefore, the GT detector must be efficient to deal with the transition effects. Further, the GT detection becomes more complex in the presence of object motions and also in case of camera motion [7]. Most of the GT detectors extract the keypoints [8] i.e. interesting isolated points from each frame and uses them for computing the similarities between the frames. In general, the spatial coordinates of the edges, corners and regions are used as keypoints [9], [10]. The algorithms such as Speed Up Robust Features (SURF) [11], Maximally Stable Extremal Region (MSER) [12], and Binary Robust Invariant Scalable Keypoints (BRISK) [13] use regions of the frame objects while the other algorithms viz. Min EigenValue (MEV) [14], Corner Detectors (CD) [15], and Features from Accelerated Segment Test (FAST) [16], [17] use the edges of the frame objects as keypoints. These algorithms suffer from blurring, illumination variation, smooth edges, etc. As a result, the performance of these algorithms in terms of accuracy is moderate.

This paper proposes a novel Hybrid Keypoint Detection (HKD) algorithm for GT Detection (GTD). This framework is based on hybrid features that consider edges and regions for extracting the salient keypoints. The proposed algorithm is tested against the video datasets containing fifty video files with gradual transitions. The experimental results show that the accuracy is much better than the popular algorithms such as MEV, CD, FAST, SURF, MSER and BRISK. Therefore, the proposed algorithm helps in overcoming the false shot identification in SBD.

The rest of the paper is organized as follows: In section 2, the related works that have been carried out in the area of video boundary detection are described. Section 3 discusses the proposed gradual transition framework. In section 4, experimental results are discussed and finally in section 5, the inference and concluding remarks are furnished.

II. Related Work

For the past two decades several research work have been carried out in the area of shot boundary detection that includes both AT and GT. This section primarily focuses on the related work that use keypoints for GT boundary detection, and also addresses a few other approaches for GT.

In general, the research work on GT can be categorized into two types viz. information based approaches and spatial keypoints based approach. The information based approaches use overall summary of visual features for boundary detection whereas keypoint based approaches extract interesting points on the edges or regions in a frame for the SBD.

The work on SBD using information based approaches applies several measures such as Mutual Information (MI), Joint Entropy (JE), frequency counts, etc. Cerenkov et al [1] has proposed shot boundary detection algorithm using Mutual Information (MI) and Joint Entropy (JE). MI is a measure of information that contains intensity and color between consecutive frames to detect abrupt cut. In case of fade transition the MI is not sufficient to identify the gradual cut. Therefore, to detect gradual transition accurately JE has been applied.

A majority of information based approaches focus on the usage of histograms. Ullasgargi et al [18] applied color histogram based technique for cut detection. Though the color based SBD algorithms detect the abrupt transition well, it increases false positives in case of camera motion and illumination changes. As a result, the performance in terms of accuracy is poor and the complex gradual transitions become difficult to be detected. To overcome the drawback in Chun-Rong et al [9] introduced Contrast Context Histogram (CCH) technique that computes the contrast values of points within a region with respect to a salient corner. The CCH differs from the normal histogram in the sense that a normal histogram is insensitive to non-uniform deformation of a region. As a result, CCH is computationally efficient and highly accurate in determining the features.

Xueming Qian et al [19] applied the Accumulated Histogram Difference (AHD) for fades and flashlight detection during gradual transition. Accumulated histogram is a mapping that counts the cumulative number of intensity observations of all the bins. As the histogram based techniques ignore the spatial distribution of luminance and color, the false alarm rate becomes high. Therefore, several research works have been carried out keypoint based approaches to overcome the above shortfalls.

Most of the keypoint based approaches extract the interesting spatial coordinates of the frame such as points on the edges or regions. These approaches are explained in the following subsections:

A. Edge based Approach

In 2005, Hun-Woo et al. [20] proposed a gradual SBD technique in which the localized edge blocks were applied. The technique divided the frames into several blocks and compared the variance of local blocks between consecutive frames for detecting shot boundary. Though it seemed to be accurate, the time complexity was high as the algorithm involves several comparisons. To overcome this limitation, Xinbo Gao et al [21] proposed a feature tracking algorithm for SBD. The tracking algorithm extracts features from a set of corner points from the current frame and they are tracked with respect to each frames using window matching technique. The drawback of this approach is that the accuracy of the SBD is poor. Punitha and Jose [22] presented a SBD algorithm in which the Eigen values of each frames are computed and the minimum Eigen value is compared to detect the shot boundary. Though this method seems to be simplified, it does not perform well in case of fades in the consecutive frames.

B. Region Based Approach

In 2007, Herbert Bay et al [11] introduced SURF algorithm that uses Hessian Matrix (HM) approximation to identify the interest points and the integral matrix process is used for fast key point identification. Though the computational complexity in terms of time is less the accuracy in the identification of keypoints is poor. To overcome this drawback Huang et al [9] presented a Local keypoint matching algorithm which uses Harris corner detector to extract the keypoints and thereby to identify the SBD of all kinds of transitions. The other work related to the SBD is the MSER algorithm [12] which identifies the connected components of objects and differentiates the foreground and background to detect the SBD. Further, Stefan et al [13] introduced BRISK algorithm, which uses a mask to extract the keypoints and improves the time complexity of keypoint detection.

In the recent past, a very few research works that apply the hybrid approaches for keypoint detection. Rong Kuan Shen et al [8] presented a hybrid SBD method by integrating a High Level Fuzzy PetriNet (HLFPN) model with keypoint matching. Though the technique performs well in terms of accuracy the recall is not consistent. Therefore, in the present scenario an efficient SBD method needs to be developed and this paper proposes another hybrid approach which integrates points of interest and regions of interest to extract the keypoints and also for efficient SBD.

III. HKD ALGORITHM

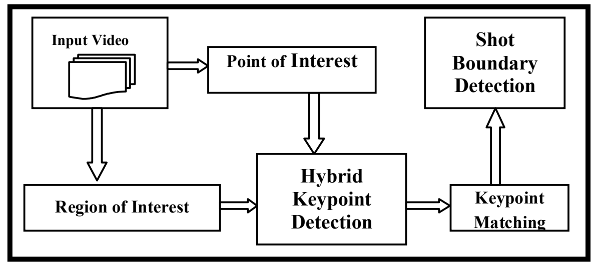

Fig . 1. HKD Algorithm

The main objective of the work is to detect gradual boundaries in the sequence of frames with higher accuracy. The block diagram of the proposed HKD framework for GTD is shown in Fig. 1. The framework receives the MPEG-4 or AVI video file of 25 frames per second as input and converts it into a sequence of video frames. The keypoints of these frames are extracted using HKD algorithm which combines the keypoints obtained using Minimum Eigen Value [23] and the second order derivatives of Hessian Matrix [11]. The salient keypoints, obtained using HKD for each frame is compared with its adjacent frame using sum of squared difference to measure the similarity. High dissimilarity distance between the frames indicates the presence of shot boundary. The HKD algorithm is illustrated in Algo. 1.

A. Hybrid Keypoints Extraction

As explained above, the HKD algorithm consists of three stages viz. keypoints extraction using Region-Of-Interest (ROI), Point-Of-Interest (POI) and keypoints matching.

Algorithm 1 Hybrid Keypoint Detection

Input Image: Gray Scale Image I_{gray}
Output Image: Keypoint based Binary Image K_{Bin}
Step1: ROI=select sub region [3 ×3] from I_{gray}
Step2: Find the Minimum Eigen Value from ROI
Step3: IM=Computation of Integral Matrix from I_{gray}
Step4: HD=Computation of Hessian Determination from IM
Step5: Interest points extract from HD
Step6: KBin =Combine features of Step2 and Step5

1. Region of Interest

The region of interest denotes a portion of image in which the required operation is performed. MEV are applied on ROI as representative values and they are useful to represent ROI in reduced format of 3x3 block. Minimum Eigen values are derived using the Eq. 1.

$$|A-\lambda I| = 0, \quad (1)$$

where A is a square matrix and Eigen values λ are the scalar values of 'A' and 'I' is the identity matrix. Among all the resultant Eigen values, positive values are considered as keypoint.

2. Point of Interest

The point of interest denotes the interesting locations which will remain stable irrespective of local and global variations with high degree of repeatability. The interesting points are identified using the Hessian Matrix (HM) which is derived from Integral Matrix (IM) [23].

IM is a cumulative sum of row and column intensities and it is calculated using Eq.2

$$\mathrm{IM}(x,y) = \sum_{i=0}^{i\le m} \sum_{j=0}^{j\le n} \mathrm{I}(i,j), \quad (2)$$

where x, y represent coordinate points of IM and i, j represent coordinate points of input image. Variables m and n represent pixel location under consideration which will be subsequently incremented till the end of the image.

Hessian is a square matrix of second order partial derivatives of a scalar valued function and it describes the local curvature of a function of many variables. The HM searches for image locations that exhibit strong derivatives in horizontal directions. The HM computes the second order derivatives L_{xx}, L_{xy}, and L_{yy} for each integral matrix point and then searches for points where the determinant of the Hessian matrix HM i.e. HD is given in Eq. 3.

$$|HM| = L_{xx}L_{yy} - (wL_{xy})^2, \quad (3)$$

where w is the weight assigned to the values of coordinate point x, y where $x<>y$ and w is used to balance the expression for the Hessian's determinant. The value of w is set as 0.9 [11]. The point of interest is extracted when the value of resultant Hessian's determinant is greater than threshold value 0.

As the keypoints provide information pertaining to their position or coverage area alone, the similarity analysis between any two consecutive frames will be inaccurate, while considering either of the two keypoints. Therefore the proposed work considers both the keypoints. As a result, there is an increase in number of keypoints and also in the accuracy.

B. Keypoint Matching

The keypoints detected using Eq. 1 and Eq. 3 yield the hybrid keypoints. For SBD, keypoint matching is essential. Keypoint matching is carried out by computing the similarity between the combined keypoints of the two consecutive frames. The frames, whose similarity distance is minimum, are considered to be the part of the same shot. Though, there are several similarity measures available in the literature, the proposed work measures the similarity, using the sum of squared difference D_1 as shown in Eq.4.

$$D1 = \sqrt{\frac{\Sigma(x-y)^2}{N-1}}, \quad (4)$$

where x, y represent current frame keypoint values and next frame keypoint values respectively. N is a number of keypoints in the frame.

IV. Experimental Results

In order to evaluate the performance of the proposed method, the experiments were carried out on fifty custom video inputs with varying contents such as animal, sea, insects, deep blue sea, birds, moving objects, human, seashore, and their combinations. The performance of the framework is measured in terms of accuracy [1] using eq. 5. Also, the accuracy of the proposed HKD algorithm is compared with the other popular algorithms viz. FAST, MEV, corner detector, SURF, BRISK, and MSER.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}, \quad (5)$$

where *TP* denotes True Positive, *TN* denotes the True Negative, *FP* denotes the False Positive and *FN* denotes the False Negative.

Table I illustrates the performance of the HKD algorithm and other benchmark algorithms in terms of average number of detected keypoints of several video inputs. The results show that the average numbers of keypoints are more in case of the HKD algorithm when compared to the other algorithms.

Table II illustrates the comparative analysis of the HKD algorithm with other benchmark algorithms based on the transitions. It is evident from the Table II that the average number of keypoints is high for HKD.

TABLE 1
COMPARATIVE ANALYSIS OF ALGORITHMS

Image ID	MEV	CD	FAST	SURF	MSER	BRISK	HFD
1	267	105	35	150	76	13	302
2	206	90	21	119	69	7	227
3	68	17	13	43	17	2	71
4	361	251	12	122	59	2	363
5	200	71	26	107	53	5	226

TABLE II
COMPARATIVE ANALYSIS ON TRANSITIONS

Transition	MEV	CD	FAST	SURF	MSER	BRISK	HKD
Abrupt	117	49	14	60	31	3	126
Gradual	221	107	21	108	55	6	238

Fig.2 shows the significance of the keypoints extracted from two consecutive input frames that are applied on the popular algorithms such as MEV, CD, FAST, SURF, MSER, and BRISK along with the proposed HKD.

The percentage of accuracy of HKD and other benchmark algorithm is shown in Table III. It is evident that the HKD algorithm outperforms all other algorithms in terms of accuracy except for the "Deep blue sea" video file in which the video contents are sea region and the algorithm fails to recognize some keypoints. In the Table III A1, A2, A3, A4, A5, A6, and A7 represents MEV, CD, FAST, SURF, MESR, BRISK, and HKD algorithms.

The performance of SBD with respect to abrupt and gradual transition is shown in Fig.3.

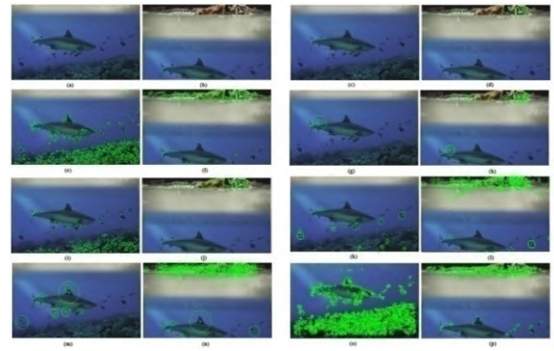

Fig. 2. Keypoint extraction for gradual transition. (a) First input frame, (b) Second input frame, (c) FAST applied on first frame, (d) FAST applied on second frame, (e) MEV applied on first frame, (f) MEV applied on second frame, (g) BRISK applied on first frame, (h) BRISK applied on second frame, (i) CD applied on first frame, (j) CD applied on second frame, (k) SURF applied on first frame, (l) SURF applied on second frame, (m) CD applied on first frame, (n) CD applied on second frame, (o) HKD applied on first frame, (p) HKD applied on second frame.

TABLE III
OVERALL ACCURACY (IN %) OF SBD

Scenes	**Region of Interest**			**Point of Interest**			**HKD**
	A1	***A2***	***A3***	***A4***	***A5***	***A6***	***A7***
Animal	95.0	70.0	82.3	94.1	80.0	23.5	98.0
Sea	53.8	28.2	7.6	92.3	56.4	5.1	97.4
Insects	92.4	80.0	96.4	92.0	80.0	78.5	96.4
Deep blue sea	13.0	4.3	4.3	4.3	13.0	4.3	60.8
Birds	89.0	60.0	14.2	85.0	73.0	14.2	95.7
Moving Objects	88.8	55.5	5.5	83.3	88.8	5.5	94.4
Animal in Water	94.5	92.6	70.5	93.0	90.0	51.4	97.0
Human	93.2	89.8	18.6	91.5	91.5	16.9	96.6
Sea shore	95.0	87.5	58.3	93.0	95.8	20.8	98.0
Human and Animals	85.0	88.0	80.0	85.0	90.0	81.2	93.7
Wild Animals	77.1	54.2	20.0	83.5	85.7	14.2	88.5

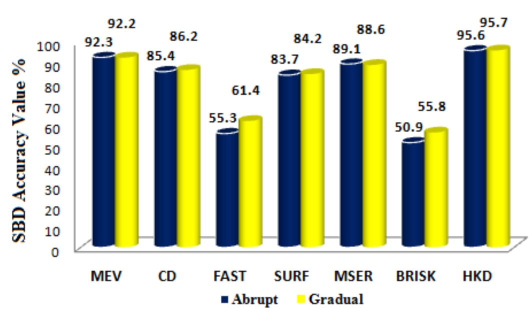

Fig. 3. Overall SBD accuracy performance (%)

V. CONCLUSION

A novel hybrid keypoint detection framework for SBD is proposed in this paper and it is experimented with several custom video inputs. The experimental results show that the HKD algorithm outperforms the other benchmark algorithms such as MEV, CD, FAST, SURF, MSER and BRISK in terms of average number of detected keypoints and accuracy. However there is a need for identifying the essential keypoints that are needed to identify the key frames. Therefore, the work can be extended to focus on the identification of minimum number of key points. Further the outcome of the HKD algorithm can be applied in the video summarization process.

REFERENCES

[1] Zuzana C erneková, Ioannis Pitas, and Christophoros "Information Theory-Based Shot Cut/Fade Detection and Video Summarization," IEEE Trans. Circuits and Systems for Video Technology, vol. 16, pp. 1051-8215, 2006.

[2] Mohanta. P. P, S. K. Saha, and B. Chanda, "A Model-based Shot Boundary Detection Technique Using Frame Transition Parameters," IEEE Trans. Multimedia, vol. 14, pp. 223-233, 2012.

[3] Joyce.R.A and Bede Liu, "Temporal Segmentation of Video Using Frame and Histogram Space,", IEEE Trans. Multimedia, vol. 8, pp. 130-140, 2006.

[4] Zhe-Ming Lu and Yong Shi, "Fast Video Shot Boundary Detection Based on SVD and Pattern Matching," IEEE Trans. Image Processing, vol. 22, pp. 5136 – 5145, 2013.

[5] Shouqun Liu, Ming Zhu and Quan Zheng, "Video Shot Boundary Detection with Local Feature Post Refinement," International Conference on Signal Processing, Beijing, pp. 1548-1551, 2008.

[6] S. Lu, Z. Wang, M. Wang, M. Ott and D. Feng, "Adaptive Reference Frame Selection for Near-Duplicate Video Shot Detection," IEEE International Conference on Image Processing, Hong Kong, pp. 2341-2344, 2010.

[7] Jeho Nam, and Ahmed H. Tewfik, "Detection of Gradual Transitions in Video Sequences Using B-Spline Interpolation," IEEE Trans. Multimedia, vol. 7, pp. 667 –679, 2005.

[8] Rong-Kuan Shen, Yi-Nan Lin, Tony Tong-Ying Juang, Victor R.L. Shen, Soo Yong Lim, "Automatic Detection of Video Shot Boundary in Social Media Using a Hybrid Approach of HLFPN and Keypoint Matching," IEEE Trans. Computational Social System, vol. 5, pp. 210 – 219, 2018.

[9] Chun-Rong Huang, Huai-Ping Lee, and Chu-Song Chen, "Shot Change Detection via Local Keypoint Matching," IEEE Trans. Multimedia, vol. 10, pp.1097 – 1108, 2008.

[10] Genliang Guan, Zhiyong Wang, Shiyang Lu, Jeremiah Da Deng, David Dagan Feng, "Keypoint Based Key frame Selection," IEEE Trans. Circuits and Systems for Video Technology, vol. 23, pp. 1 – 5, 2013.

[11] Bay, H., A. Ess, T. Tuytelaars, and L. Van Gool. "SURF: Speeded Up Robust Features," Elsevier Inc. Computer Vision and Image Understanding (CVIU), vol. 110, pp. 346–359, 2008.

[12] Obdrzalek.D, S. Basovnik, L. Mach, and A. Mikulik. "Detecting Scene Elements Using Maximally Stable Colour Regions," Springer, Communications in Computer and Information Science, vol. 82, pp. 107–115, 2010.

[13] S. Leutenegger, M. Chli and R. Y. Siegwart, "BRISK: Binary Robust Invariant Scalable Keypoints," International Conference on Computer Vision, Barcelona, pp. 2548-2555, 2011.

[14] Jianbo Shi and Tomasi, "Good Features to Track," Proc. of IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, pp. 593-600, 1994.

[15] Harris, C., and M. Stephens, "A Combined Corner and Edge Detector," Proc. of the 4th Alvey Vision Conference, University of manchestors, pp. 147-151, 1988.

[16] Biadgie.Y and K. Sohn, "Feature Detector Using Adaptive Accelerated Segment Test," International Conference on Information Science & Applications (ICISA), Seoul, pp. 1 - 4, 2014.

[17] E. Rosten and T. Drummond, "Fusing Points and Lines for High Performance Tracking," Tenth IEEE International Conference on Computer Vision (ICCV'05), Beijing, vol. 1, pp. 1508-1515, 2005.

[18] Gargi. U, R. Kasturi and S. H. Strayer, "Performance Characterization of Video-Shot-Change Detection Methods", in IEEE Trans. Circuits and Systems for Video Technology, vol. 10, pp. 1-13, 2000.

[19] Qian. X, G. Liu and R. Su, "Effective Fades and Flashlight Detection Based on Accumulating Histogram Difference," in IEEE Trans. Circuits and Systems for Video Technology, vol. 16, pp. 1245-1258, 2006.

[20] Hun-Woo Yoo & Han-Jin Ryoo & Dong-Sik Jang, "Gradual Shot Boundary Detection using Localized Edge Blocks," Multimedia Tools Appl, vol 28, pp. 283 – 300, 2006.

[21] Gao, Xinbo & Li, Jie & Shi, Yang. "A Video Shot Boundary Detection Algorithm based on Feature Tracking," Springer, International conference on Rough set and Knowledge Theory, vol. 4062, pp. 651 – 658, 2006.

[22] Punitha. P and J. M. Jose. "Shot boundary detection based on Eigen Coefficients and Small Eigen Value In Semantic," Springer, Multimedia, pp. 126 – 136, 2009.

[23] P. Viola and M. Jones, "Rapid Object Detection using a Boosted Cascade of Simple F eatures," Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Kauai, HI, USA, pp. 511-518, 2001.

Shunmugam Karpagavalli is a research scholar in the Department of Computer Science and Engineering, Manonmaniam Sundaranar University, Tirunelveli. She has received her Master of Computer Applications from St. Xaviers College, Palyamkottai, Tirunelveli. Her research interests include image and video processing.

Dr. Vanniappan Balamurugan is working as a professor in the Department of Computer Science and Engineering, Manonmaniam Sundaranar University, Tirunelveli, India. He has published more than 30 papers in the reputed journals and conferences. His research interests include Pattern Recognition, Video Processing and Natural Language Processing.

Sankaran Ramesh Kumar is a research scholar in the Department of Computer Science and Engineering, Manonmaniam Sundaranar University, Tirunelveli. He has received his Master of Engineering from Manonmaniam Sundaranar University, Tirunelveli and Bachelor of Engineering from Anna University, Chennai. His research interests include image, video processing and Pattern Recognition.