iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 **Issue:** III **Month of publication:** March 2018

DOI: http://doi.org/10.22214/ijraset.2018.3730

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue III, March 2018- Available at www.ijraset.com

Shot Boundary Detection for Gujarati News Video

Namrata Dave[1], Dr. Mehfuza Holia[2]
[1] *Gujarat Technological University*
[2]*Birla Vishwakarma Mahavidyalaya Gujarat Technological University*

Abstract: ***The method to automatically detect change in shots in video scene is called video shot boundary detection. The Shot Boundary Detection(SBD) method is widely used recently in the area of video summarization, video indexing, video data mining, etc. Usually, one or more frames from each shot are selected to represent the shot information. Such frame is termed as Key Frame. This paper presents an efficient video shot boundary detection method based on visual information-based approach which use histogram difference and rank to determine shot boundary. Experimental results indicate that the proposed method achieves accountable accuracy.***
Keywords: shot boundary detection, Key Frame Extraction, histogram difference, HSV color model, Ranking

I. INTRODUCTION

The usage of digital video is increasing on daily basis. The need for efficient processing and analysis of multimedia data has emerged on large scale. The video can be categorized in entertainment, sports, news, multimedia messages, tutorials, lectures, e-learning videos, etc. Almost all type of videos consists of information such as text, objects, shapes, textures etc. which can serve as low level features for retrieval systems. Video content can be represented by spatial, temporal or spatial-temporal characteristics. The features of video data in the spatial domain can be extracted from the frames of video, which is based on pixel information in that region and their relationship can be taken as a descriptor. The temporal domain features can be used to partition the video into frames, shots, scenes and video-segments. Any video contains two types of information: semantic content and audiovisual content. Semantic content is the main meaning conveyed by the video data. Semantic information extraction from video data can be considered as most complex task. Users can understand implicit semantic after viewing the video. Here by viewing we mean visual and audio both type of information. Users can know about explicit semantic with some intuition [1]. Textual messages present in video serve as explicit semantic. Video data normally contains audio and visual features such as color, texture, edge information, motion vectors, loudness, pitch, etc. [2, 14]. In case of textual information present in video clip, the text data which are continuously being displayed for certain time gives some important information about what is currently being viewed [3, 4]. This type of information is normally present in broadcast news video. Some of the shots in news video are having text regions which are being displayed for long duration to give idea about current topic of news, place, event or personality in news, etc. Some of the broadcasted videos contains closed caption (CC) information which is very useful for text query-based video retrieval. The close caption track is having texts to be displayed to viewers in synchronization with audio visual track. Most of the Indian news channel videos do not contain such text captions which makes retrieval task difficult [5-8].

II. RELATED WORK

A. Shot Boundary Detection

A shot can be defined as a continuous series of frames recorded in single camera action. While recording shot, the camera may stay static or may be in motion. Examples of camera motions are zooming, panning, tracking, tilting, etc. There exist a lot of content similarity between frames in a shot of the scene. Shot is considered as a basic unit of video scene. Figure 1 shows basic structure of video file. Extraction of shot from scene needs the appropriate procedure to determine the difference between two successive frames in video content. Also, it is important to determine a threshold to correctly define a shot boundary [4]. Shot changes can be categorized into sudden changes and slow changes. It is easy to detect Sudden or Abrupt change in successive frames of video. On the other hand, it is comparatively difficult to determine the gradual change in video sequence. So, one should be very careful while choosing the value of threshold as it may sometime result in false boundary if not chosen properly. There are varieties of features available for detecting shot boundary. Among available options, color features are robust in situations like complex background, orientations issues, image size variations, etc. Cut detection in case of sudden changes is easier than gradual transition detection in videos.

In literature, it is found that a lot of research has been done on shot boundary detection. The main three steps in shot boundary detection are feature extraction from a key frame, similarity detection and similarity measurement. The shot boundary detection methods use variety of features like histogram, edge, motion vectors, scale invariant feature transform (SIFT), corner points, information saliency map, etc. Each of these features is having some pros and cons. The presence of camera motion or object motion during smooth shot boundaries increases the difficulty of shot detection using color histogram features. Color histograms can be used to small camera motion. On the other hand, they are sensitive to gradual changes in scenes and cannot differentiate between the shots of a scene. So, the features using color histogram is not more useful to large camera motions. Compared to color histogram, edge features are invariant to changes in illumination as well as motion. Motion features can be very effective to handle the situation where video content is affected by motion of camera or object. However, in general color histogram can perform well on average compared to other types of features [10]. Many of the methods for shot boundary detection extracts visual features from each frame. Next step is to find frame similarities using the extracted features vectors. Based on the similarity measures shot boundaries can be detected between frames which are not similar. The Euclidean distance, the histogram intersection, etc. can be used for finding similarity for extracted feature vectors of frames. The threshold-based methods to detect short boundary are also useful in practice. The method based on threshold use predefined value of threshold. Threshold value used can be either adaptive or globally defined. Also, combination of both adaptive and global threshold serves purpose many times.

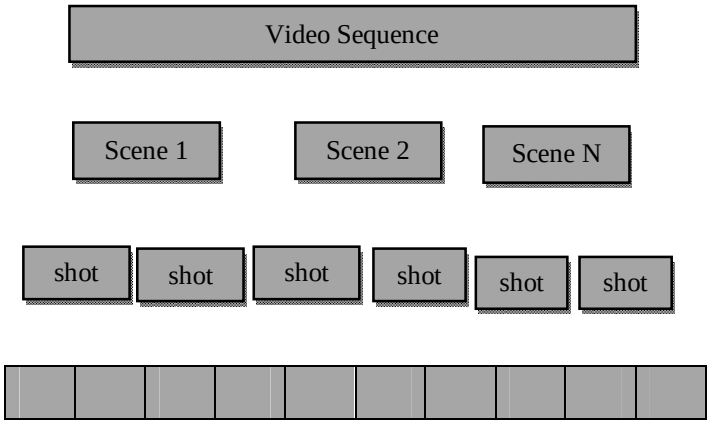


Figure 1. Video Structure

The statistical learning-based approach use classifiers to classify problem of shot boundary detection based on the features. For classification task both Supervised and unsupervised learning methods are used. The classifiers like support vector machine (SVM), Adaboost, KNN and HMM are mostly used supervised classifiers for detection of shot boundary. The unsupervised learning methods such as clustering are used to classify the frames into boundary and non- boundary. The frame similarity-based algorithms for shot boundary detection divides pairs of frames based on similarity into two clusters: one is the shot boundaries with lower values of the similarities and the other cluster of non-boundaries with higher values of the similarities. K-means, fuzzy K-means clustering, spectral clustering methods have been used for shot boundary detection [9]. Shot boundary detection approaches can be used for uncompressed domain-based and compressed domain video. The DCT features, motion vectors available in the compressed domain can be used for shot boundary detection. Many researchers have combine audio content along with visual features to improve accuracy of shot detection. The audio which is synchronized with visual part is can very useful in determining shot boundary. Approach based on perceptual pause is highly related to segmentation of temporal audio at different levels of semantic. Also finding local correlation minima is used for segmentation of audio information from given video clip. Methods based on multiple features such as anchorperson shot, caption information, voice feature and silence to segmentation of news stories from news video [10]. In video retrieval systems, one of frames is selected as key frame which represent content of video shot. Feature extraction is normally done on selected frames or key frame [12]. Features of video can be visual features, audio features, text-based features, etc. Many approaches adopted for video retrieval, summarization, and classification etc. use visual features of video frame [11]. In the key frame-based approach, shot similarities can be identified using key frames. Sometimes more than one frame can serve as key frame. Key frame is a frame chosen to represent content of entire shot. Shots of the scene are related by frame based key similarities.

Some research has been done on few Indian regional language videos for information retrieval [14]. No dataset is available to work on any of the regional language videos. Also, it has been found that due to laws of multimedia content broadcasting is different in India, due to this reason no closed caption details are available for regional language videos. Another problem is transcriptions for Indian videos are not available. Due to all such aspects, text-based approach is not found efficient [15-17].

III. PROPOSED APPROACH

In this section, we have described our proposed shot boundary detection algorithm. Our Shot boundary detection method is using HSV color model to find frame differences of successive frames of video. In the HSV color model (hue-saturation-value), hue used to represent pure colors. Saturation represents the measure of the degree where white color dilutes the pure color. HSV model is much closer to people's perception of color than RGB color model, so HSV model is used in color histogram. We are finding separate histogram for h-plane, s-plane and v-plane of color image frame extracted from input video. All three histograms are combined to represent each frame. Next step is to find histogram of all consecutive frames. At a time, algorithm will process N fix number of frames to find key frames representing the shot. Our algorithm will first try to do Matrix Factorization on Histogram vectors obtained from H, S and V histograms. Matrix Factorization will help in determining unique values from consecutive frames with help of rank of a matrix. We have described our proposed key frame extraction algorithm in figure 3. We have taken basic approach for key frame extraction. The Singular Value Decomposition theory shows that the linear transformation used is independent of scaling in each coordinate direction. The rank of a matrix can be determined as the number of linearly independent rows, which is the same as the number of linearly independent columns. So, for the diagonal matrix rank can be calculated as the number of nonzero diagonal elements. Orthogonal transforms preserve linear independence. Thus, the rank of any matrix is the

Figure 2: Key Frames determining sharp 'cut'

number of nonzero singular values. With Low Ranks of matrix, we can separate out non-similar frames. Which ultimately gives us sharp cuts in determining shot boundary of input video.

Step 1: Extract Frames from Video Clip.
Step 2: Convert each frame into HSV color space.
Step 3: Find Histogram of each of the H, S and V plane of video frame.
Step 4: Combine all three histograms to generate vector for each frame.
Step 5: Repeat Steps 2 to 4 for all frames of Video
Step 6: Apply Matrix Factorization on N number of frames
Step 7: Find rank of matrix and save it to compare it with previous ranks.
Step 8: Repeat steps 6 and step 7 to determine low ranks for given matrix

Figure 3. Key Frame Extraction Algorithm

A. Result

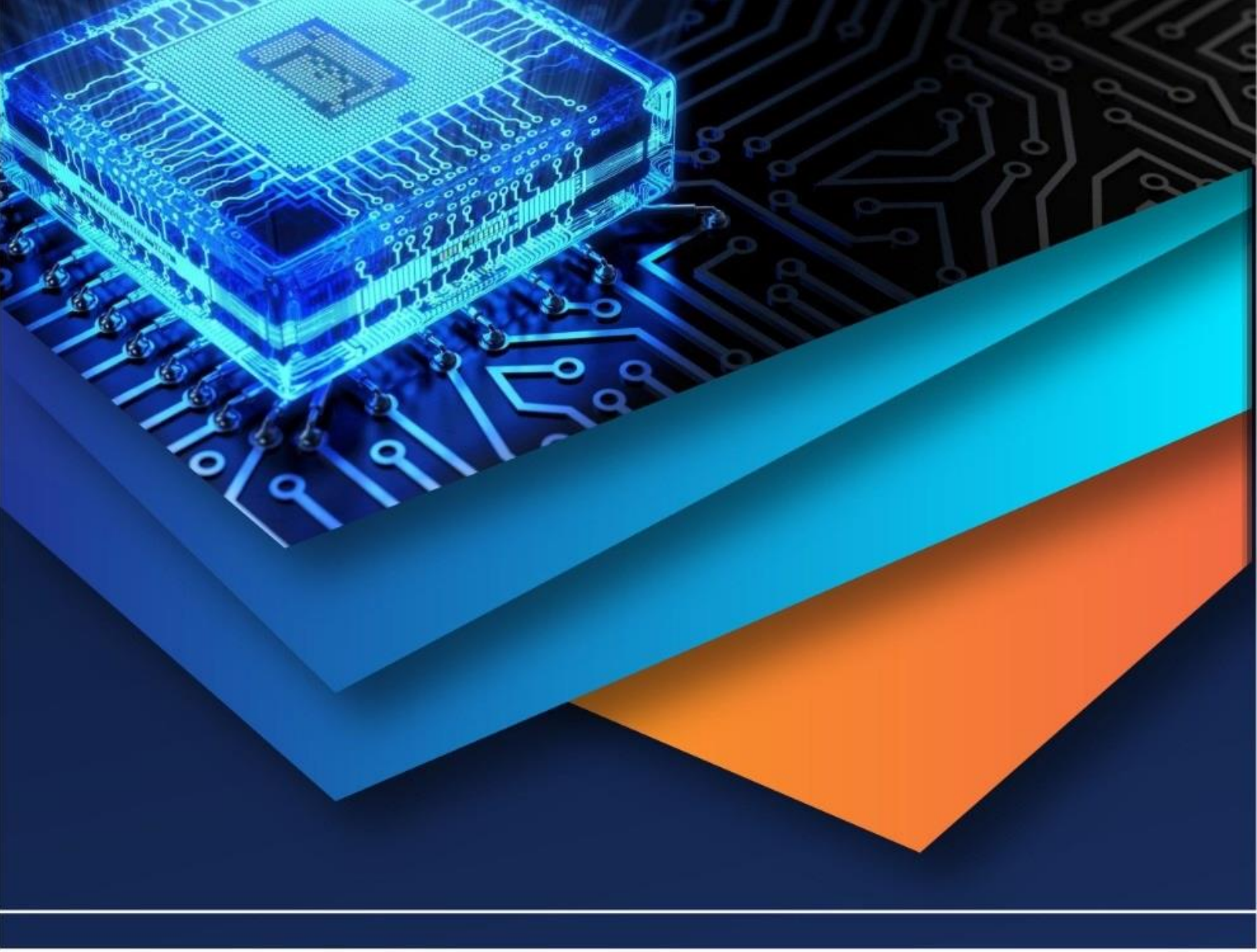
We have created our own database of News Videos of Gujarati Language. As no dataset available for Gujarati News Videos, we also have to create our own ground truth values for evaluation of our algorithm. We have evaluated this algorithm on the dataset which gives better results in determining cut in frame transitions. Output of some of the key Frames Extracted is shown in Figure 2. As it can be seen that in case of sharp cut key frames are extracted. Still we need to modify algorithm for detecting gradual transitions in shot boundary.

IV. CONCLUSION

In this paper, we have described a simpler method of extracting key frames to represent a shot in video input being read. We have applied our algorithm on our own dataset of Gujarati Language News Videos. We are able to get sharp transitions in video sequence using matrix factorization and low rank approach. In future, we further expand our algorithm to accurately find different transitions in shots such as gradual transitions, fade, dissolve and wipe.

REFERENCES

[1] Mittal, Ankush, and Sumit Gupta. "Automatic content-based retrieval and semantic classification of video content." International Journal on Digital Libraries 6.1 (2006), pp: 30-38
[2] S. Padmakala, G. S. AnandhaMala and M. Shalini, "An effective content-based video retrieval utilizing texture, color and optimal key frame features," Image Information Processing (ICIIP), 2011 International Conference on, Himachal Pradesh, 2011, pp. 1-6.doi: 10.1109/ICIIP.2011.6108864
[3] C V Jawahar, Balakrishna Chennupati, Balamanohar Paluri, Nataraj Jammalamadaka, “Video retrieval based on textual queries”, Proceedings of the Thirteenth International Conference on Advanced Computing and Communications, Coimbatore,2005.
[4] Dr P. N Chatur, R. M. Shende, “A Simple Review on Content Based Video Images Retrieval”, International Journal of Engineering Research & Technology, Volume/Issue: Vol.2 - Issue 3 (March - 2013)
[5] Tarun Jain, C.V. Jawahar, “Compressed Domain Techniques to Support Information Retrieval Applications for Broadcast Videos”, Proceedings of National Conference on Computer Vision Pattern Recognition Image Processing and Graphics (NCVPRIPG'08), pp.154-159, Jan 11-13, 2008, DA-IICT, Gandhinagar, India.
[6] Mr. Sumit R. Dhobale, Prof. Akhilesh A. Tayade , " A survey on Text Retrieval from Video " , International Journal of Application or Innovation in Engineering & Management (IJAIEM) , Volume 3, Issue 11, November 2014 , pp. 079-085 , ISSN 2319 - 4847.
[7] Rong Yan, Alexander G. Hauptmann, “A review of text and image retrieval approaches for broadcast news video”, Information Retrieval, Springer, Volume 10, Issue 4-5 , pp 445-484
[8] W. Hu, N. Xie, L. Li, X. Zeng and S. Maybank, "A Survey on Visual Content-Based Video Indexing and Retrieval," in IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 41, no. 6, pp. 797-819, Nov. 2011.doi: 10.1109/TSMCC.2011.2109710
[9] Dadashi, Roghayeh, and Hamidreza Rashidy Kanan. "AVCD-FRA: A novel solution to automatic video cut detection using fuzzy-rule-based approach." Computer Vision and Image Understanding 117.7 (2013): 807-817.
[10] Song, Yu, Wenhong Wang, and Fengjuan Guo. "News story segmentation based on audio-visual features fusion." Computer Science & Education, 2009. ICCSE'09. 4th International Conference on. IEEE, 2009.
[11] N Dave, M Holia, “Content based Video Retrieval”, IJTE special issue of ICRISET 2017, PP 156-160.
[12] Priyanka U. Patil, Dr. Krishna K. Warhade, “Analysis of Various Keyframe Extraction Methods”, International Journal of Electrical and Electronics Research, Vol. 4, Issue 2, pp: (35-40), April - June 2016
[13] Holla, K. Raghurama, and B. H. Shekar. "Video Retrieval based on Patterns of Oriented Edge Magnitude." Proceedings of the Third International Symposium on Computer Vision and the Internet. ACM, 2016.
[14] MK Audichya, JR Saini, “A Study to Recognize Printed Gujarati Characters Using Tesseract OCR”, International Journal for Research in Applied Science & Engineering, 2017.
[15] Hiranmay Ghosh, Sunil Kumar Kopparapu, Tanushyam Chattopadhyay, et al., “Multimodal Indexing of Multilingual News Video,” International Journal of Digital Multimedia Broadcasting, vol. 2010, Article ID 486487, 18 pages, 2010. doi:10.1155/2010/486487
[16] A. Jindal, A. Tiwari and H. Ghosh, "Efficient and Language Independent News Story Segmentation for Telecast News Videos," 2011 IEEE International Symposium on Multimedia, Dana Point CA, 2011, pp. 458-463.doi: 10.1109/ISM.2011.81
[17] A. Tiwari and H. Ghosh, "Ticker text extraction from Bangla news videos," 2010 Annual IEEE India Conference (INDICON), Kolkata, 2010, pp. 1-4.doi: 10.1109/INDCON.2010.57, pp12595



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY