A Scheme for Shot Detection and Video Retreival using Spatio Temporal Features

GS Naveen Kumar, VSK Reddy

Abstract: There has been a revolution in multimedia with technological advancement. Hence, Video recording has increased in leaps and bounds. Video retrieval from a huge database is cumbersome by the existing text based search since a lot of human effort is involved and the retrieval efficiency is meager as well. In view of the present challenges, video retrieval based on video content prevails over the existing conventional methods. Content implies real video information such as video features. The performance of the Content Based Video Retrieval (CBVR) depends on Feature extraction and similar features matching. Since the selection of features in the existing algorithms is not effective, the retrieval processing time is more and the efficiency is less. Combined features of color and motion have been proposed for feature extraction and Spatio-Temporal Scale Invariant Feature Transform is used for Shot Boundary Detection. Since the characteristic of color feature is visual video content and that of motion feature is temporal content, these two features are significant in effective video retrieval. The performance of the CBVR system has been evaluated on the TRECVID dataset and the retrieved videos reveal the effectiveness of proposed algorithm.

KEY WORDS: Shot Transition, Selective Frames, Integrated Feature Extraction, Feature Matching, Retrieval.

I. INTRODUCTION

With technological progress, there has been a revolution in multimedia content in the web, has resulted in many large personal and public digital video databases [1][6]. However the rapid development in the availability of the multimedia database is not accompanied by the technologies used for its efficient usage and retrieval. This is due to the fact that the content of multimedia database is not same as that of the text database which is easily accessible based on the keywords of the document [8][9]. Therefore it is very important to segment and organize the video data so as to be accessed easily. It induces the interest of researchers to focus more on the area of video processing. Although many research works have been carried out and many systems have been developed in the area of video segmentation and video retrieval, both computational cost and accuracy of the existing systems are still far beyond users' satisfaction [11]. The main issue to be addressed in the case of video segmentation is feature selection which is robust to illumination, camera and object motion and a measure of dissimilarity to detect the boundaries [12]. In this paper, an attempt has been made to address the issues involved in shot detection [13] and video retrieval which are the two important tasks of video processing.

Almost all the existing systems depend on the traditional features such as color, texture and shape feature [2] which are extracted from the segmented regions of the image. D. Saravanan et al. (2015) [19] designed a Histogram Clustering Technique in order to fast retrieve the videos from a given dataset. In Histogram Clustering Technique, the video was first divided into sequence of frames. Then, the video clustering algorithm was used in order to group the videos in different classes in which two searching was carried out. But it is very difficult to achieve reliable retrieval of the videos using single feature which results in more misdetection and false detection [14]. To overcome the above drawbacks, integration of both spatial and temporal features has to be done. As stated to the newly hypothesized algorithm by acquiring the methods like HSV color histogram [15]and motion histogram [16]to extricate color and motion features. The spatial information acquires the color features and temporal information acquires the motion feature. The arrangement of the left over sections in the manuscript is stated below: The hypothesized scheme is discussed in section II. Experimental outcomes are graphed in section III. Finally, chapter IV addresses the paper's conclusion.

II. PROPOSED SPATIO-TEMPORAL FEATURE EXTRACTION

Spatio-temporal feature extraction involves low level feature extraction and high level feature extraction as well. Temporal feature extraction is associated with motion feature extraction which depends on object movement. Therefore, spatio temporal features give an efficient and effective video retrieval.

The proposed algorithm uses two kinds of features: They are 1) Color feature 2) Motion feature

2.1 Color Feature Extraction

HSL (shade, immersion, delicacy) and HSV (tint, immersion, esteem) are elective portrayals of the RGB shading model, structured during the 1970 [17]. The HSL model endeavors to look like progressively perceptual shading models, for example, the Natural Color System (NCS) or Munsell shading framework, setting completely immersed hues around a hover at a softness estimation of 1⁄2, where a gentility estimation of 0 or 1 is completely dark or white, individually.

RGB to HSV conversion: RGB values are normalized by 255 at first. Let Vmax match the highest value of r, g, and b, and let Vmin match the lowest.

Hue computation

$$H = \begin{cases} 0, & Vmax = Vmin \\ 60° \times \left(\frac{g-b}{Vmax-Vmin} \ mod \ 6 \right), & Vmax = r \\ 60° \times \left(\frac{b-r}{Vmax-Vmin} + 2 \right), & Vmax = g \\ 60° \times \left(\frac{r-g}{Vmax-Vmin} + 4 \right), & Vmax = b \end{cases}$$

(2.1)

Saturation computation

$$S = \begin{cases} 0, & Vmax = Vmin \\ 0, & Vmax = 0 \\ \left(1 - \frac{Vmin}{Vmax} \right), & Otherwise \end{cases}$$

(2.2)

Value computation

$$V = Vmax \qquad (2.3)$$

Algorithm 1: HSV Colour space Histogram

Step 1: Extract frames from each shot/ video

Step 2: In general the frame has RGB colour space which has to be converted into HSV colour space

Step 3: The components H,S and V are evenly digitized as 16 bins for the component H, 4 bins for S and 4 bins for V

Step 4: On integrating all the components a 256 dimensional vector is extracted.

2.2 Effective Motion Feature Extraction Algorithm

In the last few years, passion moving handling has actually raised with developments moving evaluation method and also handling abilities [18]. It has numerous applications, consisting of things based video clip coding (e.g. MPEG-4), things discovery and also monitoring, vehicle driver support, scene adjustment discovery, and so on. One more typical movement evaluation issue is to acquire detailed info concerning relocating and also fixed things existing in a scene.

Algorithm 2: Motion Histogram Feature Extraction

Step 1: Segregate the frames from the shot/video
Step 2: the frames are transformed to I, P, and B frames

Step 3: P frames are extracted since it has motion information

Step 4: Obtain motion histogram with one hundred and twenty one bins

Step 5: iterate the above steps for all the p frames in the video

Step 6: obtain normalize histogram by considering the average of all the attained histograms

Step 7: do again steps 1 to 6 for all shots

2.3 CBVR System Design

The existing CBVR systems are based on low level features [10]. We can't describe the complete video content with the help of low level features since video description varies from person to person with respect to psycho visual system [20]. To design a CBVR system, first, what type of video database we are working on has to be decided? Video databases are of many types such as medical, sports, news, animation, lecture, movies and cartoon video databases. Secondly, what type of query given by user should be resolved? The types of query are text based, image based, object based, sketch based, shot based and scene based queries [21].

The objective of CBVR system is to retrieve the data base videos whose content is similar to that of user query. Direct comparison between two videos is very difficult because of huge content, a large number of computations and a lot of time consumption [22]. It is easy to compare the features of a query video to those of a database video that is to be retrieved rather than comparing of entire video content [23]. Therefore, the designers of CBVR systems have to follow two major steps- Feature Extraction and Similarity matching [24][25]. The proposed frame work is shown in figure 1.

Algorithm 3: Proposed Video Retrieval

1. A video is selected from the data base

2. The technique Spatio-Temporal Scale Invariant Feature Transform Algorithm (SIFT) [3] [7] is used to

    detect the Shot transitions.

3. Representative frames are extracted using Image Information theory

4. A hybrid feature vector is generated by the combination of HSV color histogram and motion feature vector.

5. Iterate steps 1 to 4 for rest of the videos in the data base.

6. Similarity matching is done between using Euclidean distance.

7. Videos are obtained based on less than the limit value of the range.
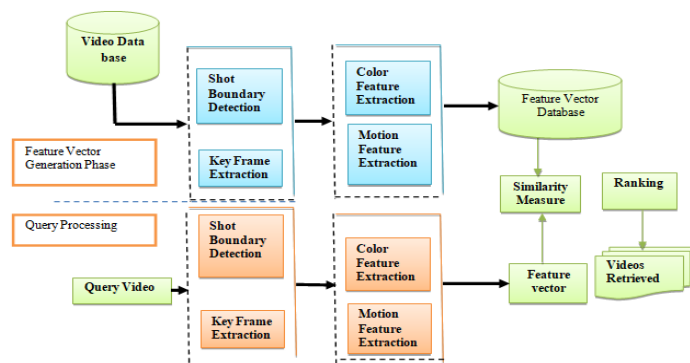

Figure 1: Proposed Framework for Video Retrieval using Spatio-Temporal feature extraction

III. EXPERIMENTAL OUTCOMES

In this research work, we took 400 videos from TRECVID data set from various categories like Entertainment,

Sports, Cartoon, News, Commercial, and Lecture. First, we integrated vectors for color and motion features [4][5]and stored them in the database for the feature vector. We followed the suit for the offline video database and web query that generated an integrated function vector. Euclidean distance was later used to calculate the similarity between the vector of the query feature and the database of the vector function. The limit value was used to determine resemblance or dissimilarity between the request and the vector function of the database. The videos are obtained from the video database that are indexed with comparable function vectors. Figure 2 shows the query video and Figure 3 shows the recovered videos.

Figure 2: Input Video

Figure 3: Retrieved Videos using Proposed Scheme

Performance Evaluation

The performance of the hypothesized Spatio-Temporal Feature Extraction algorithm Hybrid Color and Motion Histogram algorithm (HCMH) is evaluated on a database of three hours of news, commercials, documentaries, sports, movie and cartoon video sequences.

$$Recall = \frac{Number\ of\ True\ videos\ retrieved}{Total\ number\ of\ True\ Videos} \qquad (3.1)$$

Whereas Precision is defined as

$$Precision = \frac{Number\ of\ True\ Videos\ Retrieved}{Total\ number\ of\ Videos\ retrieved} \qquad (3.2)$$

F1 measureof Recall and Precision, can be defined as

$$F1\ measure = \frac{2 \times Recall \times Precision}{(Recall + Precision)} \qquad (3.3)$$

The detector is said to be good if it gives high value for both Precision and Recall. Table 1 shows Precision, Recall and F1 Measure for different genre using Existing Histogram

Clustering (HC) and Proposed Spatio-Temporal Feature Extraction method Hybrid Color and Motion histogram (HCMH) methods respectively.

Table 1: Precision, Recall and F1 measure using Existing & Proposed methods

Type of Video	Existing Histogram Clustering		Proposed HCMH Method	
	Precision	Recall	Precision	Recall
Entertainment	0.950	0.850	0.962	0.922
Sports	0.870	0.843	0.929	0.923
Cartoon	0.852	0.926	0.966	0.942
News	0.901	0.892	0.958	0.952
Commercial	0.873	0.827	0.913	0.886
Lecture	0.954	0.947	0.959	0.970
Mean value	**0.900**	**0.880**	**0.947**	**0.932**

Figure 4 and Figure 5 shows the performance comparison of Histogram Clustering (HC) and HCMH method in terms of Precision and Recall respectively for the values specified in table 1.
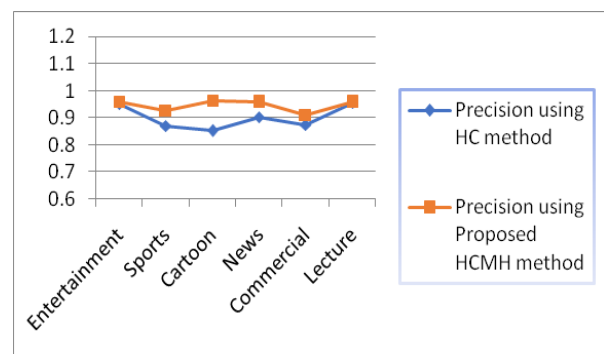

Figure 4: Performance comparison of Histogram Clustering and HCMH in terms of Precision
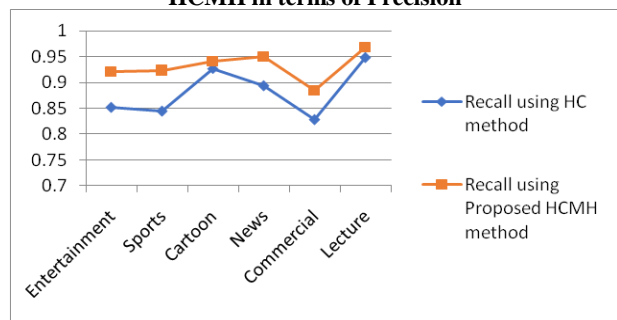

Figure 5: Performance comparison of Histogram Clustering and HCMH in terms of Recall

As shown in Figure 4 and in Figure 5, the proposed yields better result in terms of Precision, and Recall the proposed algorithm shows better result compared to Existing.

IV. CONCLUSION

In this paper, the novel spatio-temporal feature extraction algorithm Hybrid Color and Motion Histogram method is proposed. Since video is the temporal unit, a motion feature is combined with the spatial features in this proposed work. After combining the motion feature, the overall performance has been increased from 90% to 94.6% in terms of Precision, 87% to 93.1% in terms of Recall and 88% to 93.5% in terms of F1-measure as shown in performance analysis. When using TRECVID data set also the overall performance has been increased from 90% to 94.3% in terms of Precision, 89% to 93.25% in terms of Recall as shown in Table 1. The experimental result shows that the detector performs better in terms of Precision and Recall.

REFERENCES

[1]. C. Cotsaces, N. Nikolaidis, and I. Pitas, "Video Shot Detection and Condensed Representation", IEEE Signal Processing Magazine, March, 2006, pp. 28-37, 2006.

[2]. Chen, Ling, and Yuhong Wang. "Automatic key frame extraction in continuous videos from construction monitoring by using colour, texture, and gradient features." Automation in Construction (2017), Elsevier, 355-368.

[3]. D. G. Lowe, "Distinctive image features from scale-invariant keypoints," International Journal of Computer Vision, vol. 60, pp. 91-110, 2004.

[4]. Gharbi, Hana, Mohamed Massaoudi, SahbiBahroun, and EzzeddineZagrouba. "Key frames extraction based on local features for efficient video summarization." In International Conference on Advanced Concepts for Intelligent Vision Systems, pp. 275-285. Springer, Cham, 2016.

[5]. GS Naveen Kumar, and V. S. K. Reddy. "Key Frame Extraction Using Rough Set Theory for Video Retrieval." In Soft Computing and Signal Processing, pp. 751-757. Springer, Singapore, 2019.

[6]. Guru, D. S., MahamadSuhil, and P. Lolika. "A novel approach for shot boundary detection in videos." In Multimedia Processing, Communication and Computing Applications, pp. 209-220. Springer, New Delhi, 2013.

[7]. Hannane, Rachida, AbdessamadElboushaki, KarimAfdel, P. Naghabhushan, and Mohammed Javed. "An efficient method for video shot boundary detection and keyframe extraction using SIFT-point distribution histogram." International Journal of Multimedia Information Retrieval 5, no. 2 (2016), Springer, 89-104

[8]. J. H. Yuan, H. Y. Wang, and B. Zhang, "A formal study of shot boundary detection", Journal of Transactions on Circuits and Systems for Video Technology, vol. 17, no. 2, pp. 168-186, February 2007.

[9]. Kabbai, Leila, AymenAzaza, MehrezAbdellaoui, and Ali Douik. "Image matching based on lbp and sift descriptor." In 2015 IEEE 12th International Multi-Conference on Systems, Signals & Devices (SSD15), pp. 1-6. IEEE, 2015.

[10]. Li, Jun, et al. "A divide-and-rule scheme for shot boundary detection based on SIFT." JDCTA 4.3 (2010): 202-214.

[11]. Lina Sun and Yihua Zhou, "A key frame extraction method based on mutual information and image entropy," 2011 International Conference on Multimedia Technology, Hangzhou, 2011, pp. 35-38Liu, Gentao, et al. "Shot boundary detection and keyframe extraction based on scale invariant feature transform." Computer and Information Science, 2009. ICIS 2009. Eighth IEEE/ACIS International Conference on. IEEE, 2009.

[12]. Liu, Huayong, and HuifenHao. "Key frame extraction based on improved hierarchical clustering algorithm." In 2014 11th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), pp. 793-797. IEEE, 2014.

[13]. Liu, Xin, and Jin Dai. "A method of video shot-boundary detection based on grey modeling for histogram sequence." International Journal of Signal Processing, Image Processing and Pattern Recognition 9, no. 4 (2016): 265-280.

[14]. Lu, Zhe-Ming, and Yong Shi. "Fast video shot boundary detection based on SVD and pattern matching." IEEE Transactions on Image processing 22, no. 12 (2013): 5136-5145

[15]. Mohamadzadeh, Sajad, and Hassan Farsi. "Content based video retrieval based on HDWT and sparse representation." Image Analysis & Stereology 35, no. 2 (2016): 67-80.

[16]. Montazer, Gholam Ali, and DavarGiveki. "Content based image retrieval system using clustered scale invariant feature transforms." Optik 126, no. 18 (2015): 1695-1699.

[17]. Nasreen, Azra, and G. Shobha. "Reducing redundancy in videos using reference frame and clustering technique of key frame extraction." In International Conference on Circuits, Communication, Control and Computing, pp. 348-440. IEEE, 2014.

[18]. Qu, Zhong, Lidan Lin, TengfeiGao, and Yongkun Wang. "An improved keyframe extraction method based on HSV colour space." JSW 8, no. 7 (2013): 1751-1758.Ren, Liping, et al. "Key frame extraction based on information entropy and edge matching rate." Future Computer and Communication (ICFCC), 2010 2nd International Conference on. Vol. 3. IEEE, 2010.

[19]. Saravanan, D., & Vengatesh, K. J. (2015). Video content reterival using historgram clustering technique. Procedia Computer Science, 50, 560-565.

[20]. Thakre, K. S., A. M. Rajurkar, and R. R. Manthalkar. "Video partitioning and secured keyframe extraction of MPEG video." Procedia Computer Science 78 (2016), Elsevier, 790-798.

[21]. Upesh Patel, Pratik Shah, and PradipPanchal, "Shot Detection using Pixel wise Difference with Adaptive Threshold and Colour Histogram Method in Compressedand Uncompressed Video" International Journal of Computer Applications (0975- 8887), Volume 64– No.4, February 2013

[22]. Wu, Zhonglan, and Pin Xu. "Shot boundary detection in video retrieval." Electronics Information and Emergency Communication (ICEIEC), 2013 IEEE 4th International Conference on. IEEE, 2013.

[23]. Xu, Jingwei, Li Song, and RongXie. "Shot boundary detection using convolutional neural networks." In 2016 Visual Communications and Image Processing (VCIP), pp. 1-4. IEEE, 2016.

[24]. Yuan, Jinhui, et al. "A formal study of shot boundary detection." IEEE transactions on circuits and systems for video technology 17.2 (2007): 168-186.

[25]. Zhao Guang-sheng , A Novel Approach for Shot Boundary Detection and Key Frames Extraction, 2008 International Conference on Multimedia and Information Technology,IEEE.