Content Based Video Retrieval—Methods, Techniques and Applications

Reddy Mounika Bommisetty, P. Palanisamy, and Ashish Khare

Abstract Videos are rich information sources than individual images, they are considered as most influential communication media compared to others. The amount of video data produced and dispensed are growing exponentially day by day with the availability of electronic media such as smart phones, handicams etc. and broadband services at cheaper rates, as well as easy accessibility of those media in the market. Video data storage and access founds its applications in different fields such as digital libraries, video on demand, entertainment etc. and these applications are popular and needs regular access of videos from the libraries. All the above said compound reasons demanded the need of development of efficient video management and retrieval systems which can efficiently retrieve videos similar to the query as well as with a less response time. Video retrieval is made possible by searching of the desired video through a user demanded query. The user inputted query may be in the form of representative keywords or a single image or group of images. The video retrieval systems are classified as text based or content based, according to the query inputted by the user. In a text based video retrieval system query is in the form of representative keywords and the database videos are tagged with appropriate text. An example of concept based search and retrieval system is YouTube. The principal drawback in concept based system is mapping of high level or rich semantics to low level features, which is known as semantic gap. Another drawback in concept based video retrieval systems is intention gap, which denotes gap between query at querying time and intention of the search. Several researchers found content based video retrieval (CBVR) system as solution to the drawbacks of a concept based video retrieval system. The main objective this chapter is to provide comprehensive

R. M. Bommisetty (✉) · P. Palanisamy
Department of Electronics and Communication Engineering, National Institute of Technology Tiruchirappalli, Tiruchirappalli, India
e-mail: mounika@nitt.edu

P. Palanisamy
e-mail: palan@nitt.edu

A. Khare
Department of Electronics and Communication Engineering, University of Allahabad, Prayagraj, India
e-mail: ashishkhare@hotmail.com

S. Dash et al. (eds.), *Advanced Soft Computing Techniques in Data Science, IoT and Cloud Computing*, Studies in Big Data 89, https://doi.org/10.1007/978-3-030-75657-4_4

outlook on content based video retrieval (CBVR) system and its recent developments and a new content based video retrieval system that is going to be developed by feature fusion. The generalized algorithm of CBVR and its individual stages such as keyframe extraction and feature extraction also will be described elaborately. This chapter focuses on a brief overview of CBVR, keyframe extraction, feature extraction and feature fusion.

Keywords Content based video retrieval (CBVR) · Keyframe extraction · Feature extraction · Feature fusion · Semantic gap · Intention gap

1 Introduction

Video is considered as rich powerful and effective way of communication source among all the other communication resources such as text, audio, image etc. In this technological era vast availability of high quality video capturing devices at cheaper rates and internet services even in remote locations made several videos stored on one's personal devices. At one point of time manual search of desired video among the several stored videos becomes a tedious task. Furthermore, cheaper storage devices and high speed web services made lot of video data stored and shared on web. Video is quite popular since it embeds textual, visual and audio embedded in it through which we even convey our emotions along with the information. Different category videos such as sports, e-lectures, traffic, entertainment, surveillance etc. owes their own different in built characteristics. For example, traffic videos contain large movement of vehicles, e-lecture videos contains a stationary background with little bit foreground object movement. Foreground object movement is very less e-lecture videos compared to traffic videos. Based on visual information human can easily understand and interprets about the video. However, for a computer organizing or arranging such all categories of several videos and retrieving those videos later for usage is not such an easy task and involves group of several complex tasks such as spatial/temporal segmentation [1–6], feature extraction [7–10] etc. A video retrieval system is designed for storing, organizing and retrieving videos.

The goal of video retrieval system is to retrieve videos similar to the user defined query. Similarity can estimated through feature extraction and matching. Features can be low level features or high level semantics [11–13] describing the video content. Video retrieval algorithms accepts query either in form of text [11] or sample image or sequence of sample images [14]. Video retrieval algorithms based on text is popular. An example of text based retrieval system is YouTube. In a text based video retrieval system, user inputs query describing about features of video. The algorithm searches through the database and retrieves the videos which are tagged with the user described bag-of-words [15]. The principal drawback of text based retrieval system is the presence of a huge semantic gap [16] in mapping the rich user defined semantics with low level features of the video and thus results a lot of undesired content. Manually annotating all the videos in the database and designing proper keywords to describe the video content is a complex tedious task. Another drawback is intention gap [17] between the query at querying time and intention of the search.

To overcome drawbacks of text based retrieval systems researchers developed Content based Video retrieval system (CBVR) [14] which does not need any manual tagging of keywords with the database videos and retrieval accuracy also improved more compared to text based video retrieval systems. The goal of CBVR is to retrieve videos similar to the query on the basis of visual features present in the video. CBVR systems accept query in form of a sample image or video clip. CBVR extracts features of user inputted query and those features are matched with the features of database videos for retrieving videos similar to the query. Figure 1 shows general CBVR framework. CBVR systems do not require manual tagging and retrieves visually similar videos and most of the CBVR techniques employ keyframe extraction techniques [18–23] for representing video with less number of frames which reduces computational complexity and time.

The above mentioned benefits of CBVR motivated researchers to further explore in this field to improve accuracy even for complex videos.

The rest of the chapter is organized as follows. Section 2 discusses some of the past works in the domain of content based video retrieval. Section 3 discusses the proposed content based video retrieval and its individual stages along with the experimental results. Section 4 discusses few applications of the domain content based video retrieval. Finally, Sect. 5 concludes the chapter.

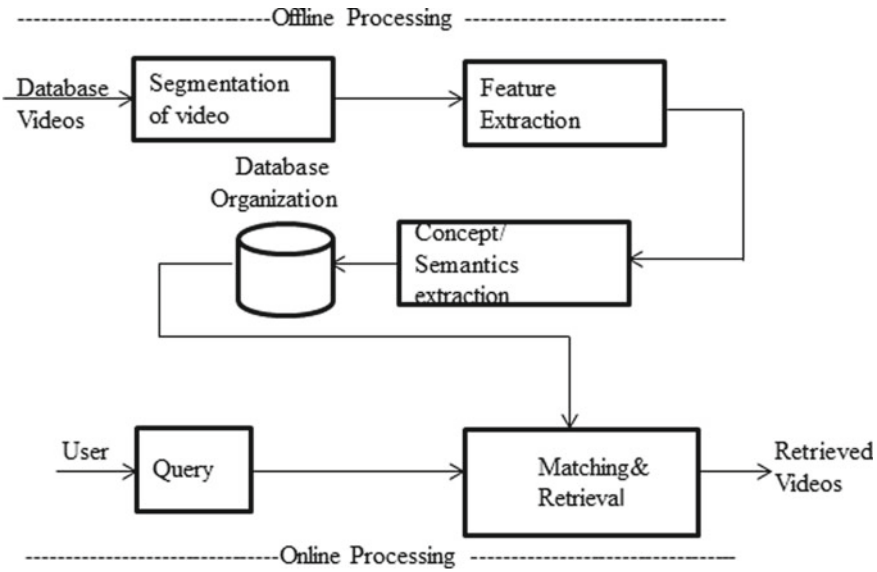

Fig. 1 Content based video retrieval system

2 State-of-the-Art Techniques

Developments in multimedia technology brought vast usage of video data as information or communication source. During recent days, smart phones equipped with high resolution camera are easily available in the market with affordable prices. Technological innovations made high speed web services available to people in cheaper rates and even accessible through smartphones with the same high speed. Several computer vision applications such as video on demand, e-learning, biomedical and surveillance applications as well as availability of several videos on internet made researchers attracted towards video management retrieval. Lokoč et al. [24] presented good review on a class of interactive video retrieval scenarios and their evaluation methods to analyse performance of new interactive video retrieval approaches. Dong et al. [25] presented video retrieval technique in which a text based query is inputted by user and the algorithm searches for unlabelled videos which are tagged with text similar to the user defined query. Their algorithm is dependent on dual deep encoding network. Wu et al. [26] proposed a unsupervised deep video hashing framework used for large scale video retrieval. They integrated video representation with optimal code learning, with provision of an efficient alternative approach to optimize the objective function. Lokoc et al. [27] presented an interactive video retrieval systems using multi-modal search and convenient inspection of results and highlighted query modification statistics. Zhang et al. [28] developed a large scale video retrieval system using a sample image query. They employed convolutional neural networks and Bag of Visual Word for representing video, and a visual weighted inverted index have been introduced by them to improve efficiency and accuracy of the retrieval process. Kordopatis-Zilos et al. [29] introduced a Fine-grained Incident Video Retrieval (FIVR) system aim of which is for a query video, the objective is to retrieve all associated videos. Their method is a single framework containing several retrieval tasks as special case. Rosetto et al. [30] presented a review and results analysis of few retrieval systems. Shen et al. [31] developed a video retrieval system using Similarity-Preserving Deep Temporal Hashing (SPDTH) and their model captures spatio-temporal properties of videos for generation of binary codes. Sauter et al. [32] vitrivr multimedia retrieval stack prepared for participating in 9th Video Browser Showdown (VBS) 2020. In this they presented extra add on facilities to the existing system such as support for classical Boolean queries, metadata filters and added a new object detection module. Zhang et al. [33] proposed a video retrieval system based on query image. They employed keyframe extraction and feature aggregation for retrieval. Sandeep et al. [34] introduced a video retrieval algorithm using hash function generated by tucker decomposition. Thomas et al. [35] proposed a synopsis based video retrieval algorithm using metadata of video such as background, moving object centroid trajectory. Araujo and Girod [36] developed video retrieval algorithm using multi features to retrieve videos from large repositories using query image. Shekar et al. [37] used Local Binary Pattern Variance (LBPV) to retrieve videos. LBPV is a variant of Local binary pattern (LBP), which is LBP augmented with local contrast variation in texture content. Mounika and Khare [38] employed super-resolution and

Histogram of Oriented Gradients (HOG) for content based video retrieval. Their algorithm applies super-resolution to database video frames at interval of 60 to eliminate any degradations in the frame and for the same frames HOG features are extracted and are used for matching with query HOG features and finally to retrieve videos similar to the query.

3 Content Based Video Retrieval (CBVR)

Content based video retrieval algorithm employs keyframe extraction techniques and several feature descriptors for the purpose of retrieval videos similar to the user inputted query. In this section we will discuss several keyframe extraction techniques and the proposed method of keyframe extraction, several feature descriptors available, the proposed method of CBVR and its performance analysis.

3.1 Keyframe Extraction

Video is a rich source of information with lot of redundancy present in it. Processing entire video for any application is complex, tedious job. Instead of processing entire video we can process few frames which are capable of together combinely representing the entire video. Group of such frames is known as keyframes. For developing an efficient keyframe extraction method, it is essential to choose distinguishing features. Statistical features are popular due to their simplicity. Statistical feature vector define statistical distribution of pixels in a frame. Keyframe extraction techniques can be categorized into four types: sequential, clustering, optimization and shot based. Sequential algorithms [39] calculate difference between features of two consecutive frames i.e. current and previous. When the difference is greater than a preset threshold then the current frame is declared as keyframe. The drawback of this algorithm is that the keyframes selected may fail to cover the entire information of video clip efficiently.

In clustering algorithms [40], frames of the video are clustered according to feature similarity using some clustering procedure and the cluster center is chosen as keyframe. Optimization based algorithms [41] select the keyframes by optimizing a desired objective function. Computational complexity and time complexity are high in optimization based algorithms. Shot detection based algorithms [18–20] group similar frames of the video into shots and then extracts keyframes from each shot.

In the present article, we propose a new method of keyframe extraction based on shot boundary detection approach which employ gradient as a statistical feature to calculate dissimilarity and threshold of shot boundary establishment is also designed using statistical parameters mean and standard deviation. The proposed method involves two stages:

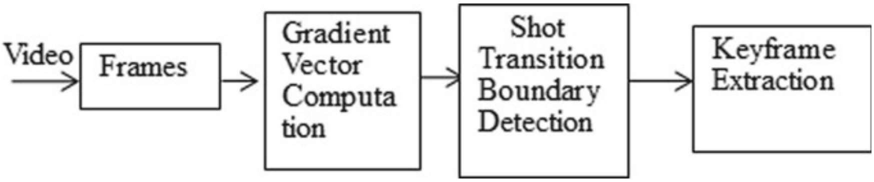
Video
Frames
Gradient Vector Computa tion
Shot Transition Boundary Detection
Keyframe Extraction

Fig. 2 The proposed method of Keyframe extraction using Gradient

i. In the first stage, shot boundaries are established by calculating gradient.
ii. In the second stage, the last frame of shots identified in the stage (i) are selected as keyframes.

The proposed methodology is attractive due to properties of its high efficiency, less computation and no supervision. The proposed methodology is based on the fact that the difference between the features of consecutive frames varies at the boundary of a shot. Block diagram of the proposed keyframe extraction method is shown in Fig. 2.

Unlike color-based feature, the gradient feature is less susceptible to local illumination changes and camera operations such as- zoom in, zoom out etc. The proposed shot detection method is described below—

Gradient magnitude and direction for each and every frame of the video are extracted. Then, for a video V comprising of N frames let GM_i, GD_i, GM_{i+1}, GD_{i+1} are the gradient magnitude and direction of two consecutive frames F_i and F_{i+1} respectively. Then, the difference in magnitude and direction of ith and $i + 1$th frame is given by the following expressions

$$DM(i, i+1) = \sqrt{\sum_{j=1}^{m}\sum_{k=1}^{n} |GM_i(j,k) - GM_{i+1}(j,k)|^2} \tag{1}$$

$$DD(i, i+1) = \sqrt{\sum_{j=1}^{m}\sum_{k=1}^{n} |GD_i(j,k) - GD_{i+1}(j,k)|^2} \tag{2}$$

where, $DM(i, i+1)$ and $DD(i, i+1)$ are dissimilarity between two consecutive frames F_i and F_{i+1} along gradient magnitude and direction respectively.

The dissimilarity in both the magnitude and direction of gradient, as above, is calculated for the entire video sequence. The threshold for establishment of shot transition boundaries is calculated as follows

$$T_M = \mu_M + (\alpha * \sigma_M) \tag{3}$$

$$T_D = \mu_D + (\alpha * \sigma_D) \tag{4}$$

where $\mathrm{T_M}$ and $\mathrm{T_D}$ are thresholds for dissimilarity matrices *DM* and *DD* respectively.

μ_M, μ_D, σ_M and σ_D are mean and standard deviations of *DM* and *DD* respectively.

α is a constant in the range [0, 1].

Now, the shot transition boundaries are established by comparing the dissimilarities with threshold values obtained using Eqs. (3) and (4). The last frame of each shot is taken as keyframe of that particular shot.

The advantages of the proposed keyframe extraction are due to the incorporation of statistical features, which have proven to be effective feature vectors, it offers ease of computation and approximately robust to noise.

3.2 Feature Vectors for CBVR

Feature descriptors describe visual properties of the video. several broad classifications of features exists such as based on the information employed global and local, Low level and High level. Low level features based on the applications used for, based on the mathematics employed for computation, statistical, Local invariant, textural and transform domain etc. Features such as shape, colour, texture etc. are low level features. High level features are used to define semantic content of the image and they depends on human visual perception. A global feature vectors constructed by taking information present in the entire image into account and global features are used for high-level applications. Features such as invariant moments, shape matrices etc. are examples of Global descriptors. Local feature descriptors are constructed from inform information present in localized portions of the image and local features are used for low-level applications. Features such as SIFT [42], SURF [43], LBP [44], BRISK [20] etc. are examples of local features. Local features offers several benefits such as—

(i) Invariance to scale, rotation, translation, illumination and occlusion.
(ii) They are distinct in nature.
(iii) Even though they captures rich information they are highly compact and efficient.

No matter, whether it is local or global if features are extracted from statistical distribution of pixels present in the frame then they can be said as statistical features. The principal advantages of statistical features are their ease of computation and approximately robust to noise. Several statistical features are available in the literature and some of them are- gradient [45], Color Moments [18], Pearson Correlation Coefficient [18], Histogram of Oriented Gradients (HOG) [38], Laplacian of Gaussian (LOG) [46] etc. Local invariant features are the features that are invariant to image rotation, scale and robust across a substantial range of affine distortion and change in illumination. Many local invariant Features such as Scale Invariant Feature Transform (SIFT) [42], Speeded Up Robust Feature (SURF) [43], Features from accelerated segment test (FAST) [47], Binary Robust Independent Elementary Features (BRIEF) [48] and Binary Robust Invariant Scalable Keypoint (BRISK) [20] etc. are developed in the past. Texture feature descriptor describes spatial arrangement of pixel intensities or colors. The texture features does not depend on the object's size, shape, orientation, and brightness. Textures may employ local information or global information. Examples of texture features that employ local information are Local Binary Pattern [44], Uniform Local Binary Pattern [19], Weber Local Binary Pattern (WLBP) [49], LBP Variance [37] etc. Examples of texture features that employ global information are Log Gabor filters [50], GLOGTH [51] etc.

The efficiency of CBVR system depends on efficient feature vectors selection and extraction. Wide number of feature descriptors are available in an image. Mostly low level features are extracted. In following we will discuss few feature vectors used in the proposed method of content based video retrieval (CBVR).

3.2.1 Laplacian of Gaussian (LOG)

To find rapid change areas in a frame Laplacian filters have been used. Noise sensitivity is more in any derivative filter. Since, Laplacian is a derivative filter, to reduce noise sensitivity, a Gaussian smoothing will be applied to the frame and then on smoothed version of frame Laplacian filtering will be applied. The process combinely is known as Laplacian of Gaussian (LOG). The LOG filter is used to highlight edges in a frame. The LOG filter detects or highlights edges in a frame through computation of second order spatial derivative. LOG yields a zero response for an image with constant intensity. The filter response will be—

(i) Zero at spatial locations far away from the edge.
(ii) Positive at darker portions located side to the edge.
(iii) Negative at brighter portions located side to the edge.

3.2.2 Histogram of Oriented Gradients (HOG)

Dalal et al. [15] introduced Histogram of Oriented Gradients (HOG). HOG descriptor gives us the number of times a particular gradient orientation appears in a localized portion of an image. The theme behind HOG is appearance and shape of a local object can be better described by the distribution of intensity gradients. The attractive feature of HOG is that it is invariant to geometric and photometric transformation. HOG efficiently captures the shape of object and foreground, structure of edge.

3.3 The Proposed Method

The block diagram of the proposed content based video retrieval algorithm is shown in Fig. 3.

In offline processing, the videos stored in database are accessed one by one. Each video of database undergo shot detection and keyframe extraction. Now the

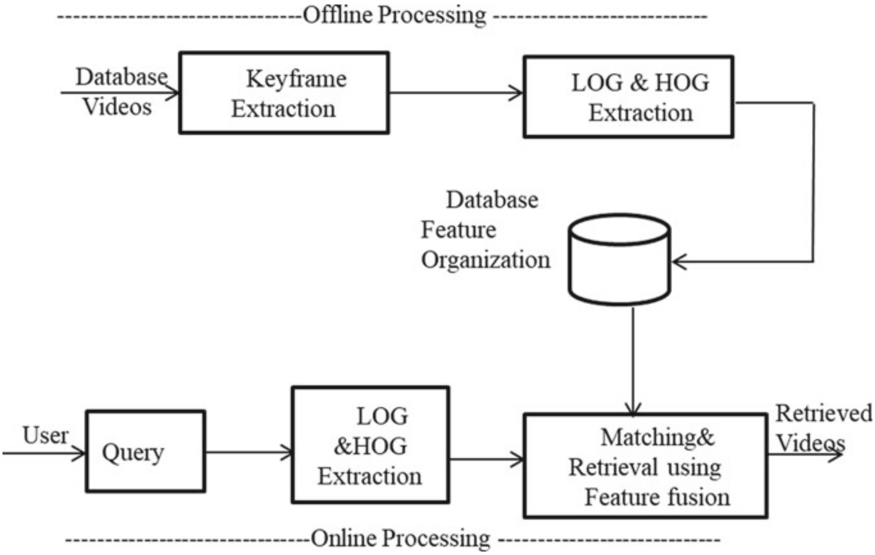

Fig. 3 The Proposed content based video retrieval system

features Laplacian of Gaussian (LOG) and Histogram of Oriented Gradients (HOG) of keyframes are extracted. The LOG and HOG features of database video keyframes together are used to represent the database videos. In online processing, user inputs query which is a sample image along with the text describing the category of the video. Then, LOG and HOG features of query image are extracted, used to represent the query image and these features are searched to find a matched features in the database. The database videos whose LOG and HOG features of keyframes matched with query frame's LOG and HOG features are retrieved separately to get final retrieval result.

The LOG and HOG features of the query frame are matched with the features of database keyframes by performing matching. For the purpose of matching we employed euclidean distance in the proposed retrieval algorithm. Let K_d= $[K_1, K_2 \ldots K_N]$ denote the keyframes of a database video V_d. Then euclidean distance is calculated between the features of query frame Q_i and features of each K_i, $i \epsilon$ $[1, N]$ of database video V_d to perform matching and its mathematical equation is given as—

$$ED = \sqrt{\sum_{i=1}^{M} \sum_{j=1}^{N} (F_q(i,j) - F_K(i,j))^2} \tag{5}$$

where,

ED Euclidean distance

$F_q(i, j)$ LOG/HOG feature of query frame q at spatial location (i, j)
$F_K(i, j)$ LOG/HOG result of keyframe K at spatial location (i, j).

Euclidean distance between both the LOG and HOG features of query frame and database video keyframes are calculated separately. Then corresponding database videos for which the euclidean distance of either LOG/HOG is less considered as matched one.

3.4 Performance Analysis

The proposed method have been experimented on the entire dataset developed by Mounika and Khare [38] which is publicly available at https://sites.google.com/site/mounikabrv3/research-profile. Quality of any proposed algorithm will be judged by analyzing performance of the method. Performance analysis can be done either qualitatively or quantitatively. Qualitative performance analysis is a subjective way of measuring quality and it depends on user and varies from user to user. Quantitative performance analysis is an objective way of measuring quality and it does not depend on user and so leads to impartial judgment. Performance of the proposed retrieval algorithm is evaluated both qualitatively and quantitatively and their results are compared with other state-of-art methods [34–38].

3.4.1 Qualitative Performance Analysis

In this section for an example query frame of dataset, top ten retrieved results extracted by the proposed method when only HOG feature used and only LOG feature used and when both the features used have been presented in Figs. 4, 5, 6 and

Fig. 4 An example query frame of serials category video of dataset

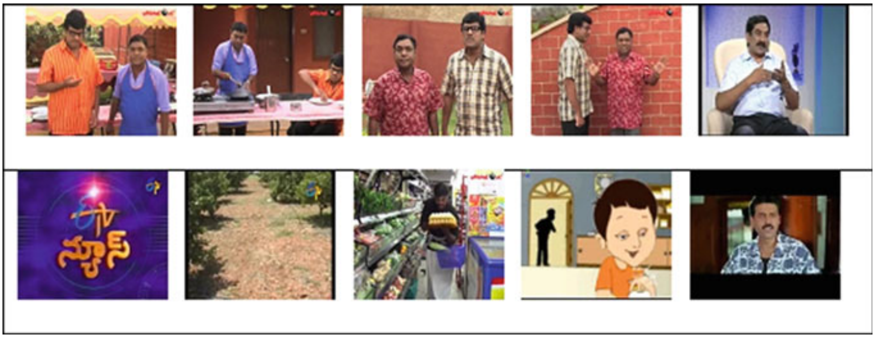

Fig. 5 Representative frames of top ten retrieved video results obtained by the proposed method when only HOG employed for query in Fig. 4

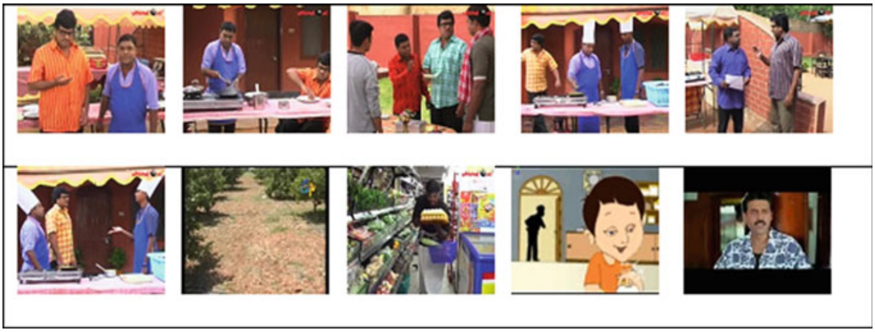

Fig. 6 Representative frames of top ten retrieved video results obtained by the proposed method when only LOG employed for query in Fig. 4

7 respectively. For the same query, top ten retrieved results extracted by the other methods the other methods [34–38] have been presented. The proposed retrieval method has been compared with five different state-of-art methods namely-Sandeep

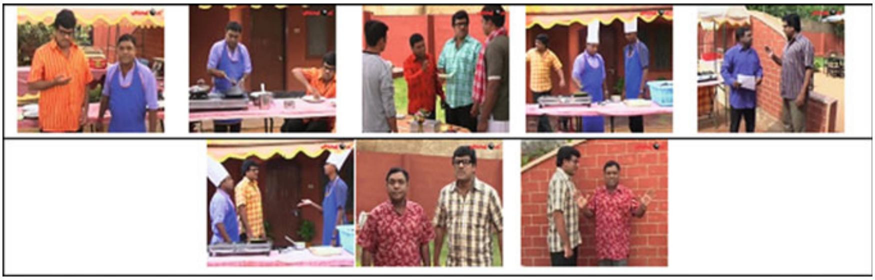

Fig. 7 Representative frames of top ten retrieved video results obtained by the proposed method using LOG and HOG feature fusion for query in Fig. 4

et al. [34], Thomas et al. [35], Araujo and Girod [36], Shekar et al. [37] and Mounika and Khare [38]. The above methods are well known, recent feature based methods of this field and all the methods employed different features, which made us to choose these methods for comparative analysis of the proposed method. The top ten retrieval results obtained by the methods- Sandeep et al. [34], Thomas et al. [35], Araujo and Girod [36] and Shekar et al. [37] and Mounika and Khare [38] have been presented in Figs. 9, 10, 11, 12 and 13 respectively.

From Figs. 5, 6 and 7, we can observe that when only HOG feature is used only 4 videos similar to query are retrieved, when only LOG feature is used only 6 videos similar to the query are retrieved and when both the features used then 8 videos similar to the query are retrieved as result. The performance is improved with fusing two features namely HOG and LOG. From Figs. 8, 9, 10, 11 and 12 we can clearly observe that the method by Sandeep et al. [34] retrieved 4 videos, the method by Thomas et al. [35] retrieved 3 videos, the method by Araujo and Girod [36] retrieved 5 videos, the method by Shekar et al. [37] retrieved 5 videos, the method by Mounika and Khare [38] which is base of the proposed retrieval method retrieved 5 videos relevant to query either in sense of object or in sense of background or even both with slight changes of illumination.

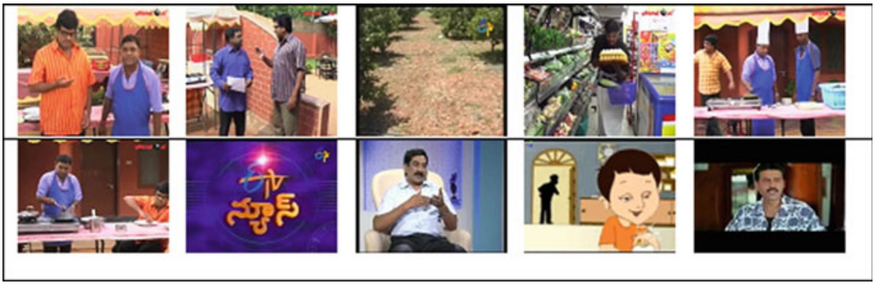

Fig. 8 Representative frames of top ten Retrieved video results obtained by Sandeep et al. [34] method for query in Fig. 4

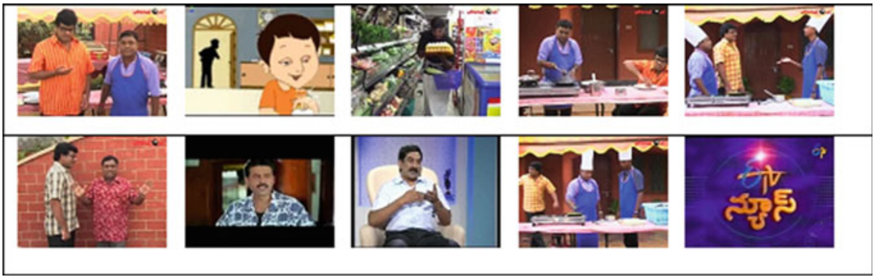

Fig. 9 Representative frames of top ten Retrieved video results obtained by Thomas et al. [35] method for query in Fig. 4

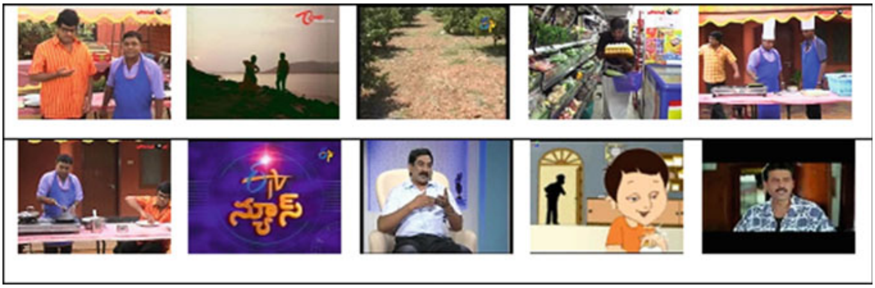

Fig. 10 Representative frames of top ten Retrieved video results obtained by Araujo and Girod [36] method for query in Fig. 4

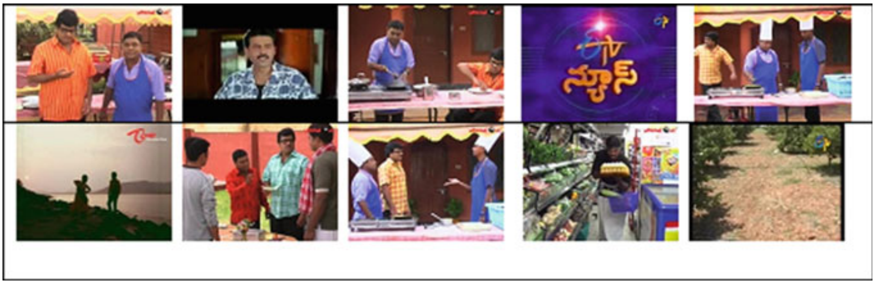

Fig. 11 Representative frames of top ten Retrieved video results obtained by Shekar et al. [37] method for query in Fig. 4

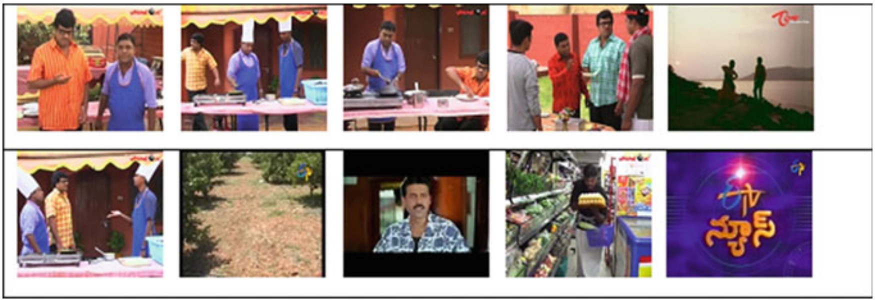

Fig. 12 Representative frames of top ten Retrieved video results obtained by Mounika and Khare [38] method for query in Fig. 4

On an overall the proposed method given good qualitative results. The benefit comes from two reasons one is fusing multiple features. In any case, if one feature fails another feature may success and the other is addition of one more field in query describing the category of the video searched for. The representative frame of the retrieved video clip in all Figs. 5, 6, 7, 8, 9, 10, 11 and 12.

3.4.2 Quantitative Performance Analysis

To analyse performance of the proposed method we chosen five different quantitative performance measures. All the chosen five performance measures are designed based on reference of ground truth. For our work, we have built ground truth manually. The proposed method performance is evaluated with five parameters they are- Precision, Recall, Jaccard index, Accuracy and Specificity. They are mathematically given as—

$$P = \frac{TP}{TP + FP} \tag{6}$$

$$R = \frac{TP}{TP + FN} \tag{7}$$

$$J = \frac{TP}{TP + FP + FN} \tag{8}$$

$$Specificity = \frac{TN}{TN + FP} \tag{9}$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{10}$$

where,

True Positive (TP) Number of videos that are identified as relevant to query by both the ground truth and the algorithm.

True Negative (TN) Number of videos that are identified as irrelevant to query by both the method and ground truth.

False Positive (FP) Number of videos that are incorrectly retrieved as relevant by the method but not present in the ground truth.

False Negative (FN) Number of videos that are relevant to query as per the ground truth but not present in the retrieved result of algorithm

From results furnished in Table 1 and Fig. 13 we can clearly observe that the proposed CBVR method with incorporation of feature fusion with HOG and LOG is

Table 1 Performance comparison of the proposed CBVR method with single feature and with feature fusion with other state-o-art CBVR methods [34–38]

Performance parameter	Only HOG	Only LOG	The proposed CBVR method
Precision	32.01	25.4	59.91
Recall	32.03	25.20	59.65
Accuracy	25.65	20.67	58.56
Jaccard index	23.35	23	82.79
Specificity	22.12	15.8	37.97

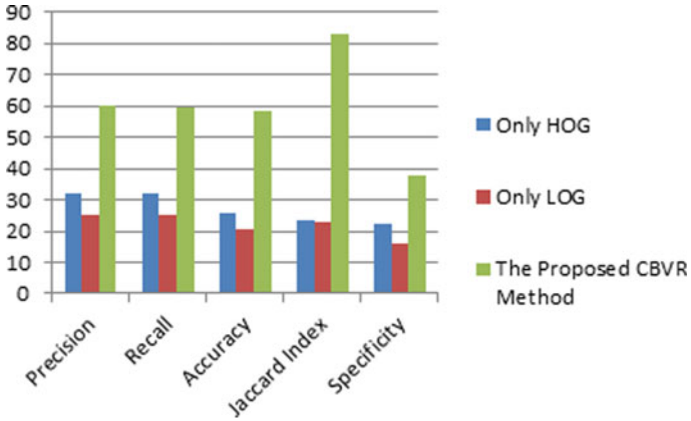

Fig. 13 Average Value of Performance Measures Versus the feature set for the proposed CBVR method

performing far better compared to incorporation of a single feature either only HOG or LOG in terms of Precision, Recall, Accuracy, Jaccard Index and Specificity.

From results furnished in Table 2, Fig. 14, we can clearly observe that the proposed CBVR method given better performance than the other methods [34–38] in terms of Precision, Recall, Accuracy, Jaccard Index and Specificity.

Table 2 Performance comparison of the proposed CBVR method with other state-of-art CBVR Methods [34–38]

Method	Precision (%)	Recall (%)	Accuracy (%)	Jaccard index (%)	Specificity (%)
Sandeep et al. [34] method	16.73	23.33	15.17	11.00	12.50
Thomas et al. [35] method	4.42	4.42	7.62	2.42	11.04
Araujo and Girod [36] method	20.83	20.83	22.22	14.61	36.10
Shekar et al. [37] method	32.65	37.07	26.91	23.68	25.44
Mounika and Khare [38] method	37.20	37.20	29.88	25.74	25.21
The proposed CBVR method	59.91	59.65	58.56	82.79	37.97

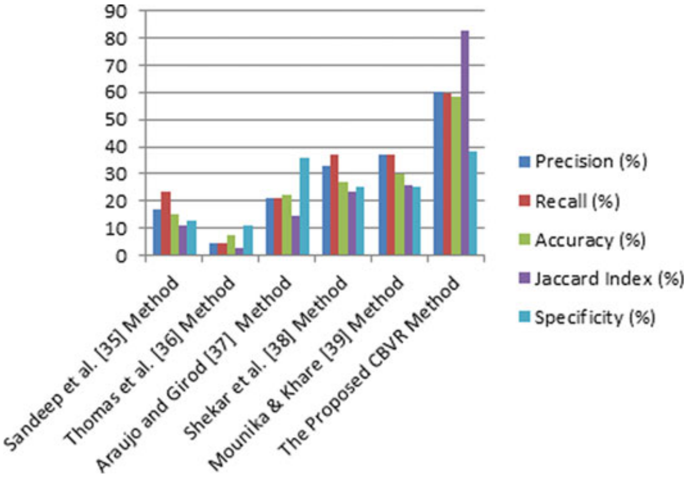

Fig. 14 Average Value of Performance Measures Versus the state-of-art methods [34–38] and the proposed CBVR method

4 Applications

Content based video retrieval founds its applications widely in several computer vision tasks such as innovative smart city applications such as surveillance, traffic monitoring, crowd monitoring, activity recognition etc. educational applications such as e-learning, bio medical applications such as monitoring different health conditions MRI Scan, surgeries to improve health care services and entertainment application such as Video-on demand etc.

5 Conclusion

This chapter discussed concept of feature fusion for content based video retrieval. Feature fusion is beneficial since, if one feature fails in capturing query content and matching it with database, another feature may success. The effectiveness of feature fusion was explained with the help of the proposed Content based video retrieval (CBVR) method. The proposed method two different features namely HOG and LOG. The proposed method's performance improved more compared to the proposed method's performance with a single feature. The application of feature fusion is not only limited to CBVR domain only. Recent advancements lead to the application of fusion concepts in several different computer vision fields. Since, different categories of videos own different properties. A single feature cannot hold the characteristics of those different videos. Hence, feature fusion finds well applicable for different video processing fields especially to CBVR and the CBVR domain founds its application in wide range of computer vision fields. Hence, the CBVR combined with feature fusion proven to be an effective research in the domain of video processing.

References

1. Mashtalir, S., Mashtalir, V.: Spatio-temporal video segmentation. In: Advances in Spatio-Temporal Segmentation of Visual, pp. 161–210. Springer, Cham (2020)
2. Xu, K., Wen, L., Li, G., Bo, L., Huang, Q.: Spatiotemporal cnn for video object segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1379–1388 (2019)
3. Tian, Y., Cheng, G., Gelernter, J., Yu, S., Song, C., Yang, B.: Joint temporal context exploitation and active learning for video segmentation. Pattern Recogn. **1**(100), (2020)
4. Jin, Y., Cheng, K., Dou, Q., Heng, P.A.: Incorporating temporal prior from motion flow for instrument segmentation in minimally invasive surgery video. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 440–448. Springer, Cham (2019)
5. Zhong, S.H., Wu, J., Jiang, J.: Video summarization via spatio-temporal deep architecture. Neurocomputing **7**(332), 224–235 (2019)
6. Zhang, K., Wang, L., Liu, D., Liu, B., Liu, Q., Li, Z.: Dual Temporal Memory Network for Efficient Video Object Segmentation (2003). arXiv preprint arXiv:2003.06125
7. Ahmadi, N., Akbarizadeh, G.: Iris tissue recognition based on GLDM feature extraction and hybrid MLPNN-ICA classifier. Neural Comput. Appl. **32**(7), 2267–2281 (2020)
8. Wu, Y.: Research on feature point extraction and matching machine learning method based on light field imaging. Neural Comput. Appl. **31**(12), 8157–8169 (2019)
9. Qin, Y., Zou, J., Tang, B., Wang, Y., Chen, H.: Transient feature extraction by the improved orthogonal matching pursuit and K-SVD algorithm with adaptive transient dictionary. IEEE Trans. Industr. Inf. **16**(1), 215–227 (2019)
10. Wang, R., Shi, Y., Cao, W.: GA-SURF: a new speeded-up robust feature extraction algorithm for multispectral images based on geometric algebra. Pattern Recogn. Lett. **1**(127), 11–17 (2019)
11. Janwe, N., Bhoyar, K.: Semantic concept based video retrieval using convolutional neural network. SN Appl. Sci. **2**(1), 80 (2020)
12. Francis, D., Anh Nguyen P, Huet B, Ngo CW. Fusion of multimodal embeddings for ad-hoc video search. In: Proceedings of the IEEE International Conference on Computer Vision Workshops, pp. 0–0 (2019)
13. Zhang, J., Mei, K., Zheng, Y., Fan, J.: Exploiting mid-level semantics for large-scale complex video classification. IEEE Trans. Multimed. **21**(10), 2518–2530 (2019)
14. Prathiba, T., Kumari, R.S.: Content based video retrieval system based on multimodal feature grouping by KFCM clustering algorithm to promote human–computer interaction. J. Ambient Intell. Humaniz. Comput. **13**, 1–5 (2020)
15. Zeng, J., Liu, M., Fu, X., Gu, R., Leng, L.: Curvature bag of words model for shape recognition. IEEE Access. **29**(7), 57163–57171 (2019)
16. Agosti, M., Marchesin, S., Silvello, G., Vezzani, F., Di Nunzio, G.M., Tellez, D., Hoppener, D., Verhoef, C., Grunhagen, D., Nierop, P., Drozdzal, M.: Learning unsupervised knowledge-enhanced representations to reduce the semantic gap in information retrieval. ACM Trans. Inf. Syst. **1**(1) (2020)
17. Song, G., Tan, X.: Deep code operation network for multi-label image retrieval. Comput. Vis. Image Underst. **1**(193), (2020)
18. Bommisetty, R.M., Prakash, O., Khare, A.: Keyframe extraction using Pearson correlation coefficient and color moments. Multimed. Syst. **18**, 1–33 (2019)

19. Mounika, B.R., Prakash, O., Khare, A.: Key frame extraction using uniform local binary pattern. In: 2018 Second International Conference on Advances in Computing, Control and Communication Technology (IAC3T), pp. 87–91. IEEE (2018)
20. Khare, A., Mounika, B.R., Khare, M.: Keyframe extraction using binary robust invariant scalable keypoint features. In: Twelfth International Conference on Machine Vision (ICMV 2019), vol. 11433, p. 1143308. International Society for Optics and Photonics (2020)
21. Mounika, B.R., Khare, A.: Shot boundary detection using second order statistics of gray level co-occurrence matrix. Res. J. Comput. Inf. Technol. Sci. **5**, 1–7 (2017)
22. Abed, R., Bahroun, S., Zagrouba, E.: KeyFrame extraction based on face quality measurement and convolutional neural network for efficient face recognition in videos. Multimedia Tools Appl. **6**, 1–22 (2020)
23. Li, Y., Kanemura, A., Asoh, H., Miyanishi, T., Kawanabe, M.: Multi-Sensor integration for key-frame extraction from first-person videos. IEEE Access. **9**(8), 122281–122291 (2020)
24. Lokoč, J., Bailer, W., Schoeffmann, K., Münzer, B., Awad, G.: On influential trends in interactive video retrieval: video browser showdown 2015–2017. IEEE Trans. Multimedia **20**(12), 3361–3376 (2018)
25. Dong, J., Li, X., Xu, C., Ji, S., He, Y., Yang, G., Wang, X.: Dual encoding for zero-example video retrieval. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 9346–9355 (2019)
26. Wu, G., Han, J., Guo, Y., Liu, L., Ding, G., Ni, Q., Shao, L.: Unsupervised deep video hashing via balanced code for large-scale video retrieval. IEEE Trans. Image Process. **28**(4), 1993–2007 (2018)
27. Lokoč, J., Kovalčík, G., Souček, T., Moravec, J., Čech, P.: VIRET: A video retrieval tool for interactive known-item search. In: Proceedings of the 2019 on International Conference on Multimedia Retrieval, pp. 177–181 (2019)
28. Zhang, C., Lin, Y., Zhu, L., Liu, A., Zhang, Z., Huang, F.: CNN-VWII: an efficient approach for large-scale video retrieval by image queries. Pattern Recogn. Lett. **15**(123), 82–88 (2019)
29. Kordopatis-Zilos, G., Papadopoulos, S., Patras, I., Kompatsiaris, I.: FIVR: Fine-grained incident video retrieval. IEEE Trans. Multimedia **21**(10), 2638–2652 (2019)
30. Rossetto, L., Gasser R., Lokoc, J., Bailer, W., Schoeffmann, K., Muenzer, B., Soucek, T., Nguyen, P.A., Bolettieri, P., Leibetseder, A., Vrochidis, S.: Interactive video retrieval in the age of deep learning-detailed evaluation of vbs 2019. IEEE Trans. Multimedia. (2020)
31. Shen, L., Hong, R., Zhang, H., Tian, X., Wang, M.: Video retrieval with similarity-preserving deep temporal hashing. ACM Trans. Multimed. Comput. Commun. Appl. (TOMM) **15**(4), 1–6
32. Sauter, L., Parian, M.A., Gasser, R., Heller, S., Rossetto, L., Schuldt, H.: Combining boolean and multimedia retrieval in vitrivr for large-scale video search. In: International Conference on Multimedia Modeling, pp. 760–765. Springer, Cham (2020)
33. Zhang, C., Hu, B., Suo, Y., Zou, Z., Ji, Y.: Large-scale video retrieval via deep local convolutional features. Adv. Multimedia. **9**, 2020 (2020)
34. Sandeep, R., Sharma, S., Thakur, M., Bora, P.K.: Perceptual video hashing based on Tucker decomposition with application to indexing and retrieval of near-identical videos. Multimedia Appl. **75**(13), 7779–7797 (2016)
35. Thomas, S.S., Gupta, S., Venkatesh, K.S.: Perceptual synoptic view-based video retrieval using metadata. SIViP **11**(3), 549–555 (2017)
36. Araujo, A., Girod, B.: Large-scale video retrieval using image queries. IEEE Trans. Circuits Syst. Video Technol. **28**(6), 1406–1420 (2018)
37. Shekar, B.H., Uma, K.P., Holla, K.R.: Video clip retrieval based on LBP variance. Procedia Comput. Sci. **1**(89), 828–835 (2016)
38. Mounika, B.R., Khare, A.: Content based video retrieval using histogram of gradients and frame fusion. In: Twelfth International Conference on Machine Vision (ICMV 2019), vol. 11433, p. 114332J. International Society for Optics and Photonics (2020)
39. Shi, Y., Yang, H., Gong, M., Liu, X., Xia, Y.: A fast and robust key frame extraction method for video copyright protection. J. Electr. Comput. Eng. (2017)

40. Kannappan, S., Liu, Y., Tiddeman, B.: DFP-ALC: automatic video summarization using distinct frame patch index and appearance based linear clustering. Pattern Recogn. Lett. **120**, 8–16 (2019)
41. Liu, X.M., Hao, A.M., Zhao, D.: Optimization-based key frame extraction for motion capture animation. Vis. Comput. **29**(1), 85–95 (2013)
42. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**(2), 91–110 (2004)
43. Bay, H., Tuytelaars, T., Van Gool, L.: Surf: speeded up robust features. In: European Conference on Computer Vision, pp. 404–417. Springer, Berlin, Heidelberg (2006)
44. Porebski, A., Vandenbroucke, N., Macaire, L.: Haralick feature extraction from LBP images for color texture classification. In: 2008 First Workshops on Image Processing Theory, Tools and Applications, pp. 1–8. IEEE (2008)
45. Liu, C.L.: Normalization-cooperated gradient feature extraction for handwritten character recognition. IEEE Trans. Pattern Anal. Mach. Intell. **29**(8), 1465–1469 (2007)
46. Yaji, G.S., Sarkar, S., Manikantan, K., Ramachandran, S.: DWT feature extraction based face recognition using intensity mapped unsharp masking and laplacian of gaussian filtering with scalar multiplier. Procedia Technol. **1**(6), 475–484 (2012)
47. Rosten, E., Drummond, T.: Machine learning for high-speed corner detection. In: European Conference on Computer Vision, pp. 430–443. Springer, Berlin, Heidelberg (2016)
48. Calonder, M., Lepetit, V., Strecha, C., Brief, F.P.: Binary robust independent elementary features. In: European Conference on Computer Vision, pp. 778–792 (2010)
49. Liu, F., Tang, Z., Tang, J.: WLBP: Weber local binary pattern for local image description. Neurocomputing **23**(120), 325–335 (2013)
50. Wang, W., Li, J., Huang, F., Feng, H.: Design and implementation of Log-Gabor filter in fingerprint image enhancement. Pattern Recogn. Lett. **29**(3), 301–308 (2008)
51. Poongothai, E., Suruliandi, A.: Global and local oriented gabor texture histogram for person re-identification. Braz. Arch. Biol. Technol. **62** (2019)