Asian Journal of Information Technology 16 (6): 486-492, 2017
ISSN: 1682-3915
© Medwell Journals, 2017

Effective Video Retrieval Using Pencil Screen Pattern Filter and Eigen based Indexing

[1]A. Packialatha and [2]A. Chandrasekar
[1]Department of Computer Science and Engineering, Sathyabama University, 119 Chennai, Tamil Nadu, India
[2]Department of Computer Science and Engineering, St. Joseph's College of Engineeing, 119 Chennai, Tamil Nadu, India

Abstract: In recent past, videos form a large part of data in the internet. Due to its ability to attract audience of all age groups, a large group of websites dedicated to browsing and watching videos have mushroomed. But regrettably, search engines do not give sufficient importance to retrieve videos. In all the video-oriented websites, the search is based on keywords typed in which produces unrelated output in terms of relevancy. To achieve content-based video retrieval, we suggest PSPF (Pencil Screened Pattern Filter). In the proposed system, the background key frame and object key frame of each and every shot of the video is obtained. Then the morphological shape of the background and objects are acquired and represented as a matrix of 0 and 1 based on grid encoding. The eigen values of these grid encoded matrices are used as metadata for indexing and retrieving videos. An index table is maintained based on the eigen values of the background key frame matrices. The index table is organized in such a way that the background key frame matrices of videos with related eigen values divide the adjacent field in the table. The eigen values of objects in the video are associated with the background key frame in the index table by means of linear chaining. Thus, we have effectively utilized the squat level characteristics of video to achieve ontological recovery of videos.

Key words: PSPF (Pencil Screen Pattern Filter), key frame formation, eigen value matrices, grid encoding, index table, object key frame

INTRODUCTION

Online servers like YouTube and Tudou show that video is no longer the realm of broadcast television only. Video has become the equidistant of choice for several people communicating via. internet and their smart phones. Videos have become one of the key means of distribution of information and its use in the academic stream is assuming enormous proportions. In addition they satisfy the requirements of those who gaze for online entertainment and this gives the explanation for the existence of a large number of websites devoted to browsing and watching videos. Wu *et al*. (2007) introduced a strategy for news video auto-documentary by mounting co-clustering to develop the duality between the stories and the textual-visual concepts (Wu *et al*., 2006).

Keywords are typed in to recover videos but much importance is not given to criteria 'relevancy'. Since, videos have great entertainment value they reach out to almost of all age groups with internet access. When it is the condition, relevancy and instant retrieval become a major need which has been failed to consider till now.

In all video-oriented websites, mining is based on the keywords the user types in. Using these keywords, the search engine will search for the matching tags accessible in the videos. Wang *et al*.(2012) describes our approach that sequentially localizes the location of video tags then implement a multiple occurrence learn approach which regards video and shot as bag and instance, respectively. The tag localization task is then described as calculating positive instances in a set of positive bags.

Each video may have many tags related to it. Here, tags refer to the idea on which the video is based on. Most of the websites allow the person who uploads the video to state their own tags. So, the tags are completely free of the website's vision. In variant websites, the words in the name of the video stated by the user will be used as tag words. None of the methods handles with the actual content of the video but it just takes words as filtering criteria for a video-based search.

The existing system exhibits the following defects: very high browsing time, since, the results produced are vast. Results are not pertinent as the tag words may be generic. There is no filtering process for unneeded videos.

Corresponding Author: A. Packialatha, Department of Computer Science and Engineering, Sathyabama University, Tamilnadu, 119 Chennai, India

MATERIALS AND METHODS

Method-system architecture: We suggest an efficient content-based video retrieval system for identifying and retrieving similar videos from a very large video database. At this point, searching the videos is based on the input as a video shot rather than the caption.

We store the video in a well organized manner so that retrieving is easier and more relevant. This is plotted by two-level indexing-first segregated on eigen values of background encoded matrix, pursued by object key frame encoded matrix.

The proposed system performs two different functions. First issue is to store videos in an efficient manner and second issue is to retrieve video exactly using indexing techniques. The process of storing videos in a database requires five complex tasks as shown in Fig. 1 namely video pre-processing, key frame generation, pencil shaded pattern extraction, grid based encoding and indexing.

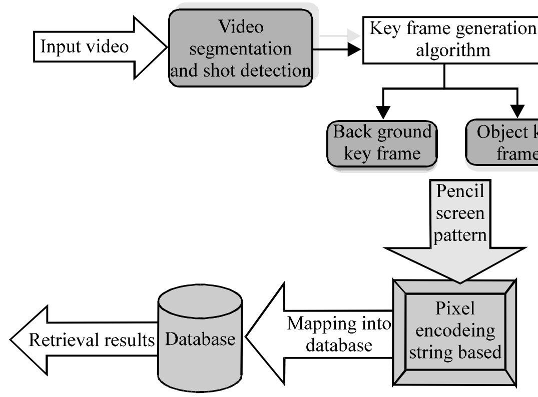

Fig. 1: System architecture diagram

Method-video processing: Video summarization can be labeled into static and dynamic summaries (Yahiaoui *et al.*, 2003; Money and Agius, 2008). In static video summary assess the content on a static storyboard with an accent on its significance or relevance where dynamic review is known as video skimming that combines video and audio information to spawn a shorter video clip (Benni *et al.*, 2015).

Knowledge cannot be gained directly from the video, since it has the most unorganised form of data by creaming video pre-processing techniques data can be predictable. snapshot act as a basic concept of a video by creaming shot revelation algorithm the video are broken down into many shots. Again shots are divided into a pack of images.

BG keyframe formation: This aspect involves in the complete formation of background frame. According to the principle of humanity the moving entity act as a foreground image and the entity that stable act as the background image primarily, the frames are counted by converting the video shots into various frames. Every frame is give as N_i [K] and its correlated with its sequential frame N_{i+1} [K] and if the frame pixel value remains the same then the frame is taken as BG key frame or else the frame is again correlated with next sequential frames. The empty pixels can be filled from the enclosing pixel (Fig. 2).

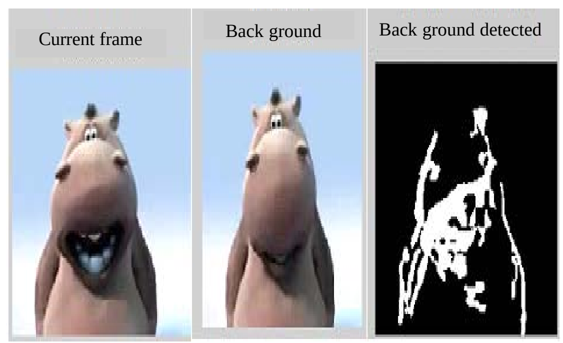

Fig. 2: Background key frame generation

Step 1: To construct the BG key frame (n_b), the following procedure has to be adopted.

Step 2: The key frames are sequentially numbered.

Step 3: Each pixel in the key frame (n_f) are compared with the same pixel in the successive frame (n_{f+1}):

- If the pixel value remains the same in both the keyframes, the pixel value is updated to the background key frame
- Else the pixel value of keyframe (n_f) is compared with other successive keyframes (n_f+2, n_f+3, ...)
- There are possibilities that few pixels may be left unoccupied in the background keyframe. The unoccupied pixel values may be computed approximately by making use of its neighbor pixels

FG (Blob) object detection: Hong *et al.* (2009, 2010), the shots displaying these sub-events are defining as key-shots which usually entail the dominant information of the result. It is pragmatic that these key-shots show regularly in the retrieved videos for a specific event query. Therefore, a scheme is projected by Hong *et al.* (2009, 2011) to generate condensed results for event-based video search. But this FG object detection aspect concerned with the abstraction of object form the key frame of video shots. The objects are abstracted from the video shot by removing the background key frame:

$$V_{seg} = \{n_i, n_{i+1}, n_{i+2}, \ldots, n_{i+x}\}$$

Step 1: Let, v_{seg} be the video segment that shares common background which is made of key frames n_i to n_{i+x}.

Step 2: To extract the visual object from the video, let us take the background key frame (n_b) and any one of the key frame in the video segment (V_{seg}). In this approach, compare the pixel of back ground key frame with any other key frame.

Step 3: During this comparison, the matching pixel are filled with the black color in the object key frame (n_o).

Step 4: The unmatched pixel color's key frame is copied into the object key frame. At the end will be able to capture the pixel occupied by the object in a video (Fig. 3).

Pencil Screened Pattern Filter (PSPF): When the background and object key frames are constructed then the key frames are screened with the black pencil by using the PSPF algorithm. Current studies on video sharing sites display that there exists a significant amount (over 30%) of duplicate videos detected in the search results (Wu *et al*., 2007).

PSPF algorithm based upon the four adjacent pixel (A_p) value is computed for the pixels. When the four pixels have the similar colour it is taken as Domestic pixel (D_p) and it left unshaded, if any of the four pixels have dissimilar colour and it is taken as Boundary pixel (B_p) and it is screened on SP_p (Screen Pattern pixels). Compare:

$$\left(A_p, 4\right) \text{in} \left(|\, n_{ij} + x \,|\right)$$
$$(\text{BG and FG frames})$$
$$\left(A_p, 4\right) \neq \left(|\, n_{ij} + x \,|\right)$$

Otherwise:

$$D_p(x) \cap SP_p$$
$$B_{p \cup SP_p}$$

Pixel encoded by grid based indexing: The object and the form of artifact structure of back ground is sized on to a grid of fixed cell in a manner that the shape is sustained to the top left corner. With the status of Web 2.0 by Hong *et al*. (2010) and Li *et al*. (2009) there emerge a few works towards the track of surveying large-scale encoded data from social sharing websites for multimedia.

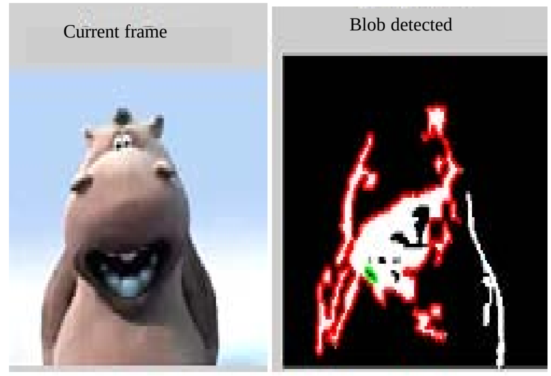

Fig. 3: Identification of object location by sorting the pixel that do not equal

The grid performed is examined from top to bottom and left to right. The 1 is assigned to cells grid wholly or partially covered by shape and cells outside of the shape boundary is given as 0 which gives sequence numbers which are used as shape representation. The disposition of binary numbers performs the framework and the object key frame. But, it most be notorious that the binary digits access for same body with a various orientation in space or with a different scale will be varied.

The pencil shaded images of background and object key frames are converted into matrices BG_k and FG_k. Eigen values are computed for both the background and object matrices using equation:

$$|\, BG_k - \lambda I \,| = 0$$
$$|\, FG_k - \lambda I \,| = 0$$

The eigen values of background and object key frames are summed up to for SE_{BG} and SE_{FG}:

$$SE_{BG} = \sum (\lambda_0 +, \ldots, + \lambda_n)$$
$$SE_{FG} = \sum (\lambda_0 +, \ldots, + \lambda_n)$$

SE_{BG} and SE_{FG} act as the metadata for the indexing of videos. Let us consider a grid matrix G of order 3×4 and the matrices of background key frame B_k and object key frame (of order 3×4) where pixel form is represented 1or 0. Pixel encoding involves the product of grid matrix G with B_k or F_k leaving the resultant matrices BG_k or FG_k. Further, the characteristic equation of BG_k and FG_k are defined and Eigen values of BG_k and FG_k are computed which are used for indexing:

$$G_r = \begin{matrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{matrix}$$

$$B_k = \begin{matrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{matrix}$$

$$BG_k = \begin{matrix} G11*B_k11 & G12*B_k12 & G13*B_k13 & G14*B_k14 \\ G21*B_k22 & G22*B_k22 & G23*B_k23 & G24*B_k24 \\ G31*B_k310 & G32*B_k32 & G33*B_k33 & G34*B_k34 \end{matrix}$$

$$FG_k = \begin{matrix} G11*F_k11 & G12*F_k12 & G13*F_k13 & G14*F_k14 \\ G21*F_k22 & G22*F_k22 & G23*F_k23 & G24*F_k24 \\ G31*F_k310 & G32*F_k32 & G33*F_k33 & G34*F_k34 \end{matrix}$$

Det (BG_k-λI) = 0 (Zhou *et al*., 2010) computes the eigen value for the back ground key frame matrix and det (FG_k-λI) = 0 computes the eigen value of the object key frame matrix.

Method-indexing in database: A database table is designed and sustained by keeping the eigen value as the major factor. Let us consider the eigen value of the BG_K frame as the key factor. The keyframes do not endure from timing and synchronization issues and can be used in a diversity of ways for browsing and navigation (Truong and Venkatesh, 2007; Money and Agius 2008; Uchiashi *et al*., 1999). For small devices, key frames may offer better browsing capabilities than video skims because they enable users to rapidly go through the relevant contents of the video in one gape fixation without watching even a small video (Wu *et al*., 2007; Money and Agius, 2008).

The structure of the table includes the fact that similar videos are placed adjacent to each other. This is done using the eigen values. Objects can be stored using linear linked list such that all the objects in relation with a particular background is linear linked.

The benefit of using our technique is the reduction in the computational time which evades the recursive computation of covariance matrix and thus raising the flexibility of choosing eigen value for k (Fig. 4).

Method-identifying similar videos: Many concerns are to be considered in order to get relevant videos proficiently and effectually. Each video must be presented more elucidative as well as in a compressed way. Chen *et al*. (2011), Packialatha and Chandra (2014), a corpus-based semantic measure was applied for calculating the similarity between videos.

Next is to scale the relevancy of videos. Considering these it gets more complicated to retrieve a video in a collection of raw videos in terms of computations. In order to set aside these issues, we suggest for a given video obtain the back ground and object key from each slot of

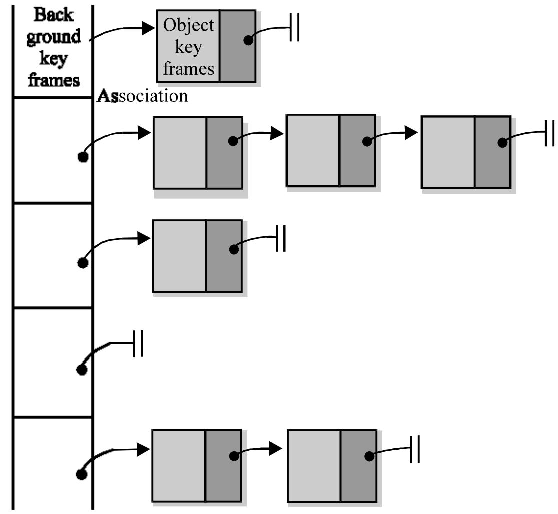

Fig. 4: Associating background and object key frame in indexing table

the morphology of these frames are taken, applied grid, converted to matrices and finally eigen values are calculated. These values are then compared with those in the table in database. Since, we use a two level comparison, i.e., both the eigen values of background and object, precision of result rate is increased to an extent. Also, relevant videos use a same eigen value is clearly proved (Zhou *et al*., 2010).

From the results any video can be selected to be viewed. One of the examples is shown, the precision of the result depends on the background, i.e., the precision rate goes down if the background is bulky or wider.

Along with the result, the notion of each related video is given. For improving and to get better result one of these concepts can be selected for filtration of the result videos. Any videos from the result can be played by the user. This act can be re done using the same or any other queries. Videos that to do not agree can be removed or resized. Hence, the complete process is easy and reasonable, quite faster (Fig. 5 and 6).

Material-practical setup: All the modules including shot detection, background subtraction, object detection are implemented using mat lab codes. The user interface is designed using forms created by mat lab codes. Eigen values are calculated using the eigen function in matlab and are stored in database. Videos are retrieved using ODBC connection with matlab. We use wamp server for simple storage and indexing. Table 1 shows the video sequences (data set). The key frames (k) of the video sequence (if, k = 6) that leads to calculate the eigen values for all video sequences. Our Pixel encoding method for calculation of eigen values λ_v is much faster than the

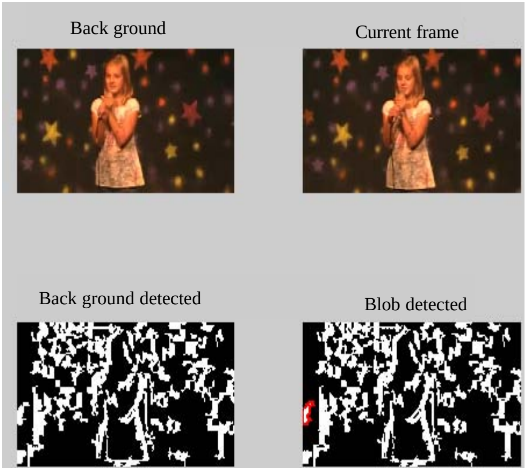

Fig. 5: Sample example of retrieval result

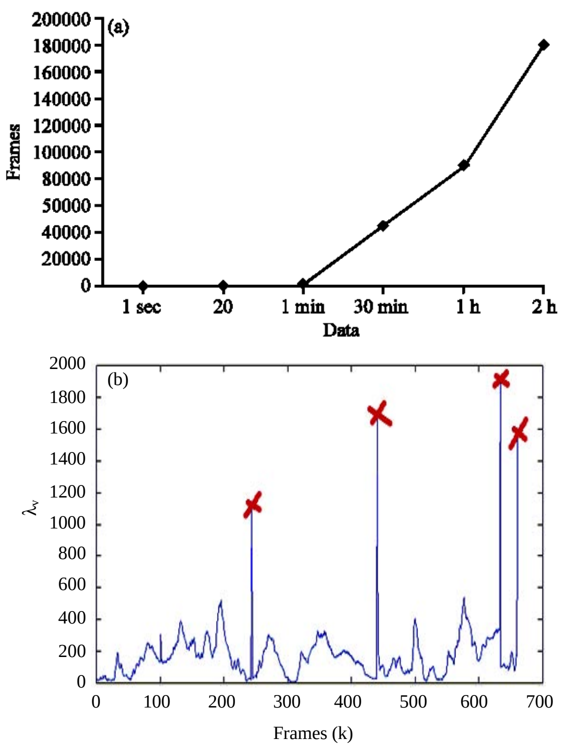

Fig. 6: a, b) Existing video retrieval system

existing method. Table 1 shows the total computational time in seconds for calculating the eigen values for the first and the second iteration.

At each iteration, the matrix is updated instead of recalculating it. Time taken from the second iteration onwards will be lesser than time taken for the first iteration. This reduces the computational time and makes advantageous over the (Packialatha and Chandra, 2014).

Table 1: Details of test videos

Data set (videos)	Frames (k)	λ_v
1 sec	25	23
20 sec	500	42
1 min	1500	112
30 min	45000	633
1 h	90000	1567
2 h	180000	2400

RESULTS AND DISCUSSION

Notable work has been done in relation to our proposal with the same objective but from a different perspective. The initiation started in 2002 when video started gaining more attention from the online users. But it was only able to provide a the oretical plan for structure examination of the images of the video for better filtration and retrieval (Ng and Lyu, 2002) and it failed to explain its practical implementation.

An enhanced research (Gitte *et al*., 2014; Huet and Merialdo, 2006) to recover videos concentrated on multimodal visual features (colour and shapes) as well, it also employed k-means algorithm for classifying the videos which suffers serious defects like the prediction of k-value, variation in clusters, etc. In additional, Shanmugam and Rajendran (2009) focussed on motion estimation and edge detection and presented them in the form of histograms. The histograms employed in visual processing are much liable to noise interference and errors.

Later, to defeat the complexity of variation in the dimension between videos, a plan came up to match low with high dimensional videos this contributed to video comparison factor (Lu *et al*., 2006). With all the technological advancements came the new plan of feature extraction for disparity of videos (Cui *et al*., 2005). Though this idea gave good similar results, it is not practical to apply it in a busy network like Internet because of its high complication and time consuming factor. Since time mattered a lot, indexing was made simple with the help of vector-based mapping which segments the videos (Varma and Talbar, 2011) later on, dynamic binary tree generation (Valdes and Martinez, 2008) came into retain memory space but it takes more time.

A similar proposal to ours came up which uses threshold and colour histogram (Varma and Talbar, 2011) to do content-based analysis but complicated in its implementation. Zheng and Zhou (2011), completely dedicated searching and retrieval method for MPEG-7 but it is not much in use in recent days. Individualized video searching with re-usability relying on the user came up with high caches (Valdes and Jose, 2011). This can become handy for private use but not for public use. When queries become tough to express, a plan came up to implement an application-based technology, combined with multi-touch exploitation which would result in enthralling the user to give entry to an external application inside their browser (Bertini *et al*., 2011).

Ultimately, the basis of our proposal is from a content-based retrieval idea (Zhou *et al*., 2010) which uses a complex B+ tree method to retrieve videos using symbolization which is practicable but complex. We have attempted to reduce the complexity level with high receptive and relevant videos with limited time utilization.

CONCLUSION

We proposed an effective CBVR. Here, we consider the morphology of the background and object in each segments of video, applied grid to get the matrix for which eigen value is calculated. This eigen values are used as metadata which will be indexed to get the videos based on the input video.

SUGGESTION

In future, we will put efforts to design an algorithm which will be more efficient to detect the background and object key frame as plain pixellize algorithm may be affected due to noise disturbance in pixels.

REFERENCES

Benni, V., R. Dinesh, P. Punitha and V. Rao, 2015. Keyframe extraction and shot boundary detection using eigen values. Intl. J. Inf. Electron. Eng., 5: 40-45.

Bertini, M., D.A. Bimbo, A. Ferracani and D. Pezzatini, 2011. Interactive video search and browsing systems. Proceedings of the 2011 9th International Workshop on Content-Based Multimedia Indexing (CBMI), June 13-15, 2011, IEEE, Madrid, Spain, ISBN:978-1- 61284-432-9, pp: 187-192.

Chen, F., D. Delannay and D.C. Vleeschouwer, 2011. An autonomous framework to produce and distribute personalized team-sport video summaries: A basketball case study. IEEE. Trans. Multimedia, 13: 1381-1394.

Cui, B., B.C. Ooi, J. Su and K.L. Tan, 2005. Indexing high dimensional data for efficient similarity search. IEEE. Trans. Knowl. Discovery Data Eng., 17: 2-15.

Gitte, M., H. Bawaskar, S. Sethi and A. Shinde, 2014. Content based video retrieval system. Intl. J. Res. Eng. Technol., 3: 430-435.

Hong, R., J. Tang, H.K. Tan, C.W. Ngo and S. Yan *et al*., 2011. Beyond search: Event-driven summarization for web videos. ACM. Trans. Multimedia Comput. Commun. Appl. TOMM., 7: 35-43.

Hong, R., J. Tang, H.K. Tan, S. Yan and C. Ngo *et al*., 2009. Event driven summarization for web videos. Proceedings of the 1st SIGMM Workshop on Social Media, October 23-23, 2009, ACM, Beijing, China, ISBN:978-1-60558-759-2, pp: 43-48.

Hong, R., J. Tang, Z.J. Zha, Z. Luo and T.S. Chua, 2010. Mediapedia: Mining web knowledge to construct multimedia encyclopedia. Proceedings of the Advances in Multimedia Modeling, January 6-8, 2010, Springer, Chongqing, China, pp: 556-566.

Huet, B. and B. Merialdo, 2006. Automatic Video Summarization. In: Interactive Video, Hammoud, R.I. (Ed.). Springer, Berlin, Germany, ISBN:978-3-540-33215-2, pp: 27-42.

Latha, P.A. and S. Chandra, 2014. Eigen based video indexing in web engine. Intl. J. Appl. Eng. Res., 9: 17051-17062.

Li, G., Z. Ming, H. Li and T.S. Chua, 2009. Video reference: Question answering on YouTube. Proceedings of the ACM 17th International Conference on Multimedia, October 19-24, 2009, ACM, Beijing, China, ISBN:978-1-60558-608-3, pp: 773-776.

Lu, H., B.C. Ooi, H.T. Shen and X. Xue, 2006. Hierarchical indexing structure for efficient similarity search in video retrieval. IEEE. Trans. Knowl. Data Eng., 18: 1544-1559.

Money, A.G. and H. Agius, 2008. Video summarisation: A conceptual framework and survey of the state of the art. J. Visual Commun. Image Represent., 19: 121-143.

Ng, C.W. and M.R. Lyu, 2002. ADVISE: Advanced digital video information segmentation engine. Ph.D Thesis, Chinese University of Hong Kong, Hong Kong.

Packialatha, A. and S.A. Chandra, 2014. Adept identification of similar videos for web-based video search. Elysium J. Eng. Res. Manage., 1: 56-60.

Shanmugam, T.N. and P. Rajendran, 2009. An enhanced content-based video retrieval system based on query clip. Intl. J. Res. Rev. Appl. Sci., 1: 236-253.

Truong, B.T. and S. Venkatesh, 2007. Video abstraction: A systematic review and classification. ACM. Trans. Multimedia Comput. Commun. Appl., 3: 1-37.

Uchihashi, S., J. Foote, A. Girgensohn and J. Boreczky, 1999. Video manga: Generating semantically meaningful video summaries. Proceedings of the 7th ACM International Conference on Multimedia (Part 1), October 30-November 05, 1999, ACM, Orlando, Florida, USA., ISBN:1-58113-151-8, pp: 383-392.

Valdes, V. and J.M. Martinez, 2008. Binary tree based on-line video summarization. Proceedings of the 2nd ACM Workshop on TRECVid Video Summarization, October 31-31, 2008, ACM, Vancouver, British Columbia, Canada, ISBN:978-1-60558-309-9, pp: 134-138.

Valdes, V. and M.M. Jose, 2011. Efficient video summarization and retrieval. Proceedings of the 9th International Workshop on Content-Based Multimedia Indexing, June 13-15, 2011, IEEE, Madrid, Spain, ISBN:978-1-61284-432-9, pp: 43-48.

Varma, S.L. and S.N. Talbar, 2011. Dynamic threshold in clip analysis and retrieval. Intl. J. Image Processi., 5: 417-424.

Wang, M., R. Hong, G. Li, Z.J. Zha and S. Yan *et al*., 2012. Event driven web video summarization by tag localization and key-shot identification. IEEE. Trans. Multimedia, 14: 975-985.

Wu, X., A.G. Hauptmann and C.W. Ngo, 2007. Practical elimination of near-duplicates from web video search. Proceedings of the 15th ACM International Conference on Multimedia, September 25-29, 2007, ACM, Augsburg, Germany, ISBN:978-1-59593-702-5, pp: 218-227.

Wu, X., C.W. Ngo and Q. Li, 2006. Threading and auto documenting news videos. IEEE. Sginal Process. Mag., 23: 59-68.

Yahiaoui, I., B. Merialdo and B. Huet, 2003. Comparison of multiepisode video summarization algorithms. EURASIP. J. Appl. Signal Process., 2003: 48-55.

Zheng, Q. and Z. Zhou, 2011. An MPEG-7 compatible video retrieval system with support for semantic queries. Proceedings of the 2011 International Conference on Consumer Electronics, Communications and Networks (CECNet), April 16-18, 2011, IEEE, XianNing, China, ISBN:978-1-61284-458-9, pp: 1035-1041.

Zhou, X., X. Zhou, L. Chen, Y. Shu and A. Bouguettaya *et al*., 2010. Adaptive subspace symbolization for content-based video detection. IEEE. Trans. Knowl. Data Eng., 22: 1372-1387.