IRACST - International Journal of Computer Science and Information Technology & Security (IJCSITS), ISSN: 2249-9555
Vol.6, No4, July-August 2016

A comparative study of content based search and retrieval of video sequences

Azra Nasreen,
Assistant Professor, Dept of CSE,
R V College of Engineering, Bangalore
Karnataka, India

Dr. Shobha G
Head of the Department, Dept of CSE,
R V College of Engineering, Bangalore
Karnataka, India

***Abstract*—This paper investigates about the search and retrieval of video sequences based on the contents. Retrieval of video sequences is an important aspect in video processing technology. This paper lays out effective approaches for searching and retrieving the relevant videos from the video databases along with benefits offered and limitations of each technique. Also, the most efficient technique with its salient features such as content based video retrieval using multiple features is discussed and also it illustrates the performance rise in such systems with the use of graphics processing unit and high performance computing.**

Keywords- content based video retrieval, feature extraction, key frames, video search, GPU, high performance computing, CUDA

I. INTRODUCTION

Emergence of social media and cheap storage devices has given rise to huge increase in sharing of videos. The amount of video content available on the web that has been shared is enormous compared to last few decades. According to a survey YouTube users alone upload 72 hours of new video content every minute. With so much of video data available, there comes a need of effective video processing techniques that can help user to browse, search, retrieve and analyze video sequences. Searching of videos is a challenging task as video contain huge amount of information and also due to the gap that exist between the low level features through which and the videos are stored and processed and high level semantics of the videos expected by the user. Due to these factors video search has become a major bottleneck for scientific research. Much of the work has been undertaken to search videos, but the existing systems are either based on text annotations, or there is a lack of optimal feature set combination to represent video data efficiently, and also inappropriate selection of feature extraction technique for generation of feature vectors results in retrieving irrelevant videos. It is still more difficult in specially challenging environments such as underwater videos , as they are seriously distorted in terms of the clarity and color fidelity comparable to the ground based tasks making the searching process much more difficult. In addition to this, as the amount of data that is processed is huge sequential processing becomes cumbersome. Multi-core architectures can be exploited by making use of graphics processing units which is more suitable for computationally intensive tasks. This paper discusses about the various aspects of CBVR such as key frame extraction, feature extraction, search and retrieval techniques in the subsequent sections.

II. KEY FRAME EXTRACTION

Video retrieval is of paramount importance in analyzing and searching the content of video sequences and is a laborious task that is computationally complex. Searching of videos can be done in two ways: with key frame extraction or without key frame extraction. Key frame extraction when carried out will reduce the quantity of data before actual search is carried out. Hence it is more feasible to have this which reduces the computational complexity of the search and retrieval tasks. Key frames are the characteristic frames of the video which render limited, but meaningful information about the content of the video. A method proposed in [1] aims to reduce animations mesh complexity by key frame extraction by using genetic algorithm to search keyframes. Binary encoding is used and the frames with highest fitness will be designated as key frames. Another method proposed in [2] identifies image epitome and Kullback Liebler Divergence (KLD) is used to measure the dissimilarity between frames. The scores are analyzed using a min-max approach to select key frames. In the above techniques, number of extracted key frames is restricted. In [3] Artificial Fish Swarm Algorithm (AFSA) is applied along with K means clustering. Color feature vector is extracted based on which clusters are formed. Middle frame from each cluster is picked as key frame. Clustering result is optimized by K-mean. Only color feature is used. In [4] Changqing Cao et al. proposed a method in which each frame is divided into blocks and the color differences in each block of the frames are used for comparison with the double thresholds. With the features of MPEG compressed video stream, an improved histogram matching method is used for video segmentation. Key frames are extracted utilizing the features of I-frame, P-frame and B-frame for each sub-lens. Fidelity and compression ratio are

used to measure the validity of the method. A real-time algorithm for scene change detection and key-frame extraction that generates the frame difference metrics by analyzing statistics of the macro-block features extracted from the MPEG compressed stream is introduced in [5]. The key-frame extraction method is implemented using difference metrics curve simplification by discrete contour evolution algorithm. An efficient approach for extracting the key frames for video abstraction and navigation called as iFrames is presented in [6]. The video is first mapped to a polyline in a high dimensional space. Then a simplification based approach is used to generate the multi-level representation of the polyline. The nodes of the polylines represent the key frames of the video. The proposed approach is complex and seeks for more efficient algorithm for projecting the frames into the high dimensional space without loss of accuracy.

III. SEARCH AND RETRIEVAL

Content analysis of videos requires the identification and extraction of low level features of videos that can be then used for searching. Content Based Video Retrieval Systems [7] [20] is an important technology that facilitates efficient search and retrieval of video sequences from the large databases on the basis of features that are extracted from the videos. The extracted features are used to index, classify and retrieve desired and relevant videos while filtering out undesired ones. Videos can be represented by their audio, texts, faces and objects in their frames. An individual video possesses unique motion features, color histograms, motion histograms, text features, audio features, features extracted from faces and objects existing in its frames. Videos containing useful information and occupying significant space in the databases are under-utilized unless CBVR systems capable of retrieving desired videos by sharply selecting relevant videos while filtering out undesired videos.

The search and retrieval method in [8] portrays the data set using Video Segmentation and Optimal Key Frame. The feature vectors are generated from VSR using texture analysis using Local Binary Pattern (LBP) technique and for color moments YCbCr model is used. Then based on the calculation of probability of occurrence of pixel intensity values with pixel location among frames, optical frames are obtained. Similarity check is done using Weighted Distance Measure. An optimal key frame representation scheme for video segmentation is proposed [9] uses video shot detection and retrieval considering both spatial and global color statistics of the objects in the frames. A technique called Temporally Maximum Occurrence Frame (TMOF) is proposed as a new method for video shot representation. Later at each pixel position of a video shot, k pixel values with highest probability of occurrence are considered to enhance the performance. The resultant methods are named as k-TMOF and k-pTMOF. Finally a comparison is made between these new schemes. The experimental results prove that the methods achieve better performance of retrieval. A method [10] for automated video indexing and video search in large lecture video database is proposed, wherein automatic video segmentation and key-frame detection is done. Textual metadata is extracted by OCR on key-frames. Automatic speech recognition is used on lecture videos. This method is dependent on the text contents of the video. Another method to gain performance over retrieving the videos using intermediate block truncation coding technique [11] divides each frame into blocks based on dimensions row wise and column wise. Then Block Truncation Coding is used for feature extraction. For each block seven different color spaces are applied to calculate the pixel intensity values. From each block the average or mean is calculated upon all seven color spaces and each color index is used to get the final vector. The same step is carried out for all other blocks and all other frames. Later all feature vectors are stored in a database. Then query feature vectors are compared with these feature vectors using absolute difference. Same method is followed [12] but for odd and even videos. An approach called mesh- local binary pattern (LBP) applied for 2D LBP variants adopted for 3D texture patterns of triangular mesh surfaces is proposed in [13]. A method for addressing rotation invariance is also given which confirms the repeatability factor. The work offers effectiveness, generalization, adaptability and simplicity factors. The experimental results prove the uniformity for various types of scalar functions and allow extension on all variants. Techniques for texture feature includes statistical analysis and structural analysis are discussed [14] and comparative analysis is made on each of the technique. There are many existing techniques and mechanisms to perform video retrieval on content basis. The default text based and color based video retrieval techniques and also techniques involving multiple features such as color and texture, texture color and shape are proposed. Techniques for shape feature include shot boundary detection and edge boundary detection and for non-visual features: captions, annotations, relational attributes and structural descriptions can be considered. There are many existing techniques and mechanisms to perform video retrieval on content basis. However, most of them are unable to perform well resulting in high ranking of irrelevant videos and also long response time leading to poor user satisfaction.

The accuracy of a video retrieval system can be greatly enhanced by having an optimal and multiple features set combination and appropriate key frame extraction as well as feature extraction technique based on the features selected. Since searching videos and retrieval is a computationally complex task, GPU can be used to exploit the multi-core architecture to extract information more efficiently. High speed-up can be achieved by distributing the tasks among multiple cores and processing them simultaneously. Challenges such as minimizing the shared data, message passing and synchronization have to be addressed while parallelizing the sequential algorithms. CUDA (Compute Unified Device Architecture) is a parallel computing environment and a programming pattern developed by NVIDIA [21]. The CUDA programming architecture consists of the components namely, the Main memory, the Central Processing Unit (CPU), the Graphics Processing Unit (GPU) and the GPU Memory. Using these components of CUDA architecture, parallel program can be developed to achieve high performance in terms of

execution time as well as accurate results using optimal feature set and appropriate feature extraction techniques to compose feature vectors. Further, content based video retrieval systems can evaluated by determining the precision and recall using (1) and (2) respectively. Video retrieval system can be measured based on the following metrics:

- **Accuracy of retrieval**: it is the ratio of count of exact matching videos that are retrieved to the total number of videos multiplied by 100.
- **Precision**: it is the ratio of count of videos retrieved similar to the query clip to the total count of retrieved videos.
- **Recall**: it is the ratio of count of retrieved videos similar to the query clip to the total count of similar videos available in the database.

$$\text{Precision} = \frac{\text{Number of retrieved videos that are relevant to query clip}}{\text{Total number of retrieved videos}} \quad (1)$$

$$\text{Recall} = \frac{\text{Total number of available videos that are relevant to query}}{\text{Number of retrieved videos relevant to the query clip}} \quad (2)$$

The metrics precision-recall calculations show the effectiveness of the CBVR system and can be used to evaluate the performance achieved.

IV. CONCLUSION

This paper proposes a comparative study of the algorithms used for key frame extraction and content based video retrieval along with the issues in the existing algorithms and the benefits offered by one over the other approaches. Content based video retrieval can be carried out by performing a pre-processing stage such as key frame extraction, followed by extraction of multiple features from these key frames and also the query and finally similarity matching of the extracted features of the query and the videos in the database. The videos with the least distance of the query will be given as the output to the user. Finally, it also brings out the advantages offered by GPU through high performance computing to the computationally complex and resource intensive tasks such as search and retrieval of videos.

REFERENCES

[1] Tong-Yee Lee, Chao-Hung Lin, Yu-Shuen Wang, and Tai-Guang Chen, "Animation Key-Frame Extraction and Simplification Using Deformation Analysis", IEEE Transactions On Circuits And Systems For Video Technology, VOL. 18, NO. 4, April 2008, pp. 478-486.

[2] C T Dang, Kumar, Radha, "*Key Frame Extraction From Consumer Videos Using Epitome*", 19th IEEE International Conference on image Processing (ICIP), Sept. 30 2012-Oct. 3 2012, pp 93 – 96.

[3] Sun Shumin, Zhang Jianming, Liu Haiyan, "Key Frame Extraction Based on Artificial Fish Swarm Algorithm and K-means", IEEE International Conference on Transportation, Mechanical, and Electrical Engineering (TMEE), December 16-18, 2011, Changchun, China, pp 1650-1653.

[4] Guozhu Liu, and Junming Zhao "*Key Frame Extraction from MPEG Video Stream*", Proceeding of the Second Symposium International Computer Science & Computational Technology (ISCSCT '09), 26-28, Dec. 2009, pp. 007-011.

[5] Changqing Cao, Zehua Chen, Gang Xie, Shaoshuai Lei, "Key Frame Extraction Based On Frame Blocks Differential Accumulation", 24th IEEE Chinese Control and Decision Conference (CCDC'12), 2012, pp 3621-3625.

[6] Yong Jin, Ligang Liu, Qingbiao Wu, "*iFrames: A Multi-Level Keyframe Extraction and Navigation Tool for Videos*", IEEE international Conference on CAD/Graphics, Sept 5-17 2011, pp 176 – 182.

[7] P. Geetha and Vasumathi Narayanan, "A Survey of Content-Based Video Retrieval", Journal of Computer Science, Vol. 4, NO. 6, pp: 474-486, 2008.

[8] S.Padmakala, Dr.G.S.AnandhaMala, M.Shalini,"An Effective Content Based Video Retrieval Utilizing Texture, Color and Optimal Key Frame Features", 2011 International Conference on Image Processing, 978-1-61284-861-7/11, 2011 IEEE.

[9] Kin-Wai Sze, Kin-Man Lam, and GuopingQiu, "A New Key Frame Representation for Video Segment Retrieval", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 15, No. 9, September 2005.

[10] Haojin Yang, Christoph Meinel, "Content Based Lecture Video Retrieval using Speech and Video Text Information", IEEE Transactions on Learning Technologies, Vol. 7, No. 2, April-June 2014.

[11] Dr. Sudeep D Thepade, Krishnasagar S Subhedarpage, Ankur A Mali and Tushar S Vaidya,"Performance Gain of Content Based Video Retrieval Technique using Intermediate Block Truncation Coding on Different Color Spaces", International conference on Communication and Signal Processing, April 3-5, 2013, India.

[12] Dr. Sudeep D Thepade, Krishnasagar, S Subhedarpage and Ankur A Mali,"Performance Rise in Content Based Video Retrieval using Multi-level Thepade's sorted ternary Block Truncation Coding with intermediate block videos and even-odd videos", International Conference on Advances in Computing, Communications and Informatics, 2013.

[13] Naoufel Werghi, Stefano Berretti, and Alberto del Bimbo, "The Mesh-LBP: A Framework for Extracting Local Binary Patterns From Discrete Manifolds", IEEE Transactions on Image Processing, Vol. 24, No. 1, January 2015.

[14] Madhav Gitte, Harshal Bawaskar, Sourabh Sethi, Ajinkya Shinde, "Content Based Video Retrieval System", International Journal of Research in Engineering and Technology, Vol. 3, Issue 6, No 2, Junne2014.

[15] Ojala, T., Pietik¨ainen, M., Harwood, D., "A comparative study of texture measures with classification based on feature distributions", Pattern Recognition, vol. 29, no. 1, pp. 51–59, 1996.

[16] B V Pateland B BMeshram, "Content Based Video Retrieval Systems", International Journal of UbiComp (IJU), Vol.3, No.2, April 2012.

[17] Hamdy K. Elminir and Mohamed Abu ElSoud, "Multi feature content based video retrieval using high level semantic concept",IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 4, No 2, July 2012.

[18] Aasif Ansari and Muzammil H Mohammed, "Content based Video Retrieval Systems - Methods, Techniques, Trends and Challenges", International Journal of Computer Applications (0975 – 8887) Volume 112 – No. 7, February 2015.

[19] NianhuaXie, Li Li, XianglinZeng, and Stephen Maybank, "A Survey on Visual Content-Based Video Indexing and Retrieval", IEEE Transactions On Systems, Man, And Cybernetics—Part C:Applications And Reviews, Vol. 41, No. 6, November 2011.

[20] M. Petkovic, W. Jonker, "Content-Based Video Retrieval by IntegratingSpatio- Temporal and Stochastic Recognition of Events", In proceedingsof IEEE Intl. Workshop on Detection and Recognition of Events inVideo, pp: 75-82, 2001.

[21] Ashok Ghatol, "Implementation of parallel image processing using NVIDIA GPU framework", advances in computing communication and control, Springer Berlin Heidelberg 2011, 457-464.