Content based video retrieval using discrete cosine transform

Sumaya Hamad[1], Ahmeed Suliman Farhan[2], Doaa Yaseen Khudhur[3]
[1,3]College of Computer Science and Information Technology, University of Anbar, Anbar, Iraq
[2]Computer Center, University of Anbar, Anbar, Iraq

Article Info

Article history:

Received Apr 18, 2020
Revised Jun 15, 2020
Accepted Jun 21, 2020

Keywords:

Content based video retrieval
Descrete cosine transform
Video retrieval
Video retrieval process

ABSTRACT

A content based video retrieval (CBVR) framework is built in this paper. One of the essential features of video retrieval process and CBVR is a color value. The discrete cosine transform (DCT) is used to extract a query video features to compare with the video features stored in our database. Average result of 0.6475 was obtained by using the DCT after implementing it to the database we created and collected, and on all categories. This technique was applied on our database of video, check 100 database videos, 5 videos in each category.

Corresponding Author:

Sumaya Hamad
Department of Computer Science
College of Computer Science and Information Technology
Anbar University, Anbar, Iraq
Email: sumayah.hamad@uoanbar.edu.iq

1. INTRODUCTION

Digital data seems to be an important part of our lives. Digital data involves videos, images, audios, documents, etc. Videos constitute an ample source of information. Video may also include all other digital content, such as photos, voices, and texts. It is further distinguished by its temporal consistency. Digital devices' fast development is causing inflation in the video store. The retrieval of the necessary information from the video database named a video retrieval process according to user needs. Video retrieval is a branched field called information retrieval. Data retrieval is considered a sub-field of computer science that arranges and retrieves data from large sets of databases. Videos retrieval approaches are important and necessary for multimedia applications like video search engines, digital archives, visual-on-demand broadcasting, etc. [1-5].

Today, the amount of open multimedia data is growing significantly as information technology and multimedia strategies develop rapidly. In major applications like monitoring, entertainment, healthcare, learning, and sports, video is heavily consumed. Searching for the elements required in these large data on the Internet can be considered an important challenge. Therefore; numerous video retrieval systems have been introduced for this intent. Video retrieval has two approaches: text-based structure, and content-based frameworks. Text descriptor is used in text-based structure to manually annotate video. The search in this method relies on the relevant metadata for. file, such as tags, titles, description and keywords. The downside of this method to manual annotation includes humanitarian job. Content-based structure is solving the above drawback. The content-based architecture allows the application to access a video clip from a list of visual content-based videos that are entirely automatically extracted, such as color, texture, shapes, and not attributes that are unrelated to the content. Visual form holds little to no semantic video content [6-11].

First, the video is split into frames, then frames are broken up into images. By image segmentation the object is isolated from the image. A part of an image is the segmented object. Extract feature from the segmented image (object) [12]. Extracted features from videos could be indexed and searched. Video search by identifying it as one of the evaluation tasks [13].

Content Based Video Retrieval (CBVR) recognized as most of the strongest functional approaches for enhanced quality of retrieval. Because of the use of wealthy material in the video, there is massive scope in the field of video retrieval to boost traditional search engine efficiency. This leads the CBVR field in a direction promising future development of more successful video search engines. Strong and expansive range of CBVR and CBVR systems is presented in a simple and detailed manner. The processes are represented in a systematic way at various stages of CBVR systems. It also displays varieties of features, their variations, and approaches, techniques, and algorithms for their use. Different querying methods, some features such as GLCM, Gabor Magnitude, similarity algorithm such as Kullback-Leibler distance method and Relevance Feedback method are addressed [14]. Implementing automatic indexing of videos, video search in a broad video database then displaying tailored results as well. The Suggested system operates in three separate steps, with video segmentation in the first phase, key frame detection retrieve relevant key frames. Second, for extracting text keyword, OCR, HOG and ASR algorithms are used over the keyframe. Color, texture and edge features will also be extracted in the third step. Eventually, analysis of the search similarity is carried out on the extracted features that are stored in the SQL database and the result is provided to the users with customized re-ranked results according to interest [15]. Video retrieval machine learning system, composed of shot extraction module, key frame extraction module and video retrieval module are also provided. One or more key-frames can then summarize each shot. Video viewing and retrieval can also be considered an issue of classification. In reality, their retrieval module is based on kernel-based SVM (Support Vector Machine), which is now represented in the active learning method, uses the extracted key frames [16].

After two years, a groundbreaking approach for achieving significantly high-quality context-based video retrieval is proposed by identifying temporal trends in video content. Based on the temporal patterns discovered, an effective indexing strategy and an efficient technique for matching sequences are combined reducing computing expenses and improve the precision of the retrieval respectively. An experimental finding show that their method is pretty good in terms of efficiency and effectiveness in the improvement of content-based video retrieval [17]. A method transfers through database video to scenes using scene change detection algorithm based on color histogram and extract key frames is suggested. Using straight forward rules multiple features are obtained for key frames. Then use the multiple feature vectors to compare query and database videos by measuring Euclidean distance [18]. Also, a system model for large video data processing in CBVR systems is presented, Usage of a distributed Computer infrastructure, with a MapReduce-based Hadoop system. They have also opted for El Ouadrhiri et al. CBVR framework as a case in point for their approach. The challenge in the current scenario is therefore the optimization of the BCMH computation time [19]. Then an automated video content analysis and retrieval system to enable the discovery of GDR TV videos in historical collections is provided. It is based on service-oriented distributed architecture and includes algorithms for shot boundary detection video analysis, idea classification, personal recognition, text recognition and similarity search [20].

Meanwhile, a video retrieval system for visual content Since a user sends a video search query in a natural language is developed. A general pipeline of a content-based video recovery system comprises two main sections : a) The offline training process in which vectors are extracted from a broad video database to train the system of video features and the corresponding video captions that describe the natural language video context, and b) Online content processing steps in which query features are picked up and then used to retrieve videos related to the database [21]. A specific video segmentation using detection of change of scene accompanied by ranking is proposed. The approach suggested will consist of the following phases: primarily Separate scenes are identified from videos to define scene boundaries. Detected boundaries help to divide videos into chunks of various scenes and then classify them. The video classifier classifies the type of a particular chunk of each frame. The video classifier output for a given chunk predicts class of highest probability. Then, eventually, user query is matched with each chunk label to obtain the appropriate piece of news [22]. By integrating multi-modal features for binary hash learning in both offline training and online query phases, multi-modal hashing methods may allow successful multimedia retrieval. Even so, when only one or some of the modalities are given, current multi-modal approaches could n’t binarize queries. Therefore, to resolve this issue, a new flexible multi-modal hashing (FMH) method is suggested. Within a single model, FMH learns various modality-specific hash codes and multi-modal cooperative hash codes concurrently. Based on the newly arriving queries, that offering any one or combination of modality characteristics, the hash codes are flexibly created. In addition, the hashing learning method is effectively monitored by the pair-wise semantic matrix to boost the ability to differentiate [23]. A series of experiments measuring how well the benchmark outcomes represent the real progress in solving the task of moment retrieval is presented. The findings show major biases in the common datasets and unexpected actions of the state-of-the-art models. In addition, the authors present new studies and strategies for visualizing the effects of health checks. At last, they recommend potential ways of improving the moment of retrieval in the future. [24]. More detailed research review can be found in [25].

The main target of this work is to build a CBVR system that supports querying by example to retrieve similar video from a database according to their features. Color features have been used to identify and describe the contents of the video. The rest of the paper is structured according to the following: Section 2 presents related research and datasets. Section 3 introduces the CBVR structure. Section 4 describes DCT feature extraction, including frame extraction and resizing, discrete cosine transformation (DCT). Section 5 reports on the test material. Section 6 presents results and analysis of the experimental study on the dataset. Section 7 concludes the paper.

2. RESEARCH METHOD

2.1. CBVR structure

The proposed framework comprises four main modules, as shown in the Figure 1. Every module has its own specific functions. The four modules are:

a) Feature Extraction module: The color features are extracted from each video and listed in the database.
b) Querying module: Through the GUI of this module retrieve the most similar video listed in the database, the user may issue his/her query.
c) Similarity matching module: Within this module, using Euclidian distance and mean square error similarity metrics, similarity between query video and database video is assessed. The best similar video will be shown according to their similarity degree.
d) Retrieval module: The database video with higher similarity rank will be retrieved and displayed to the user.

The above mention modules will be discussed in details in the next sections.

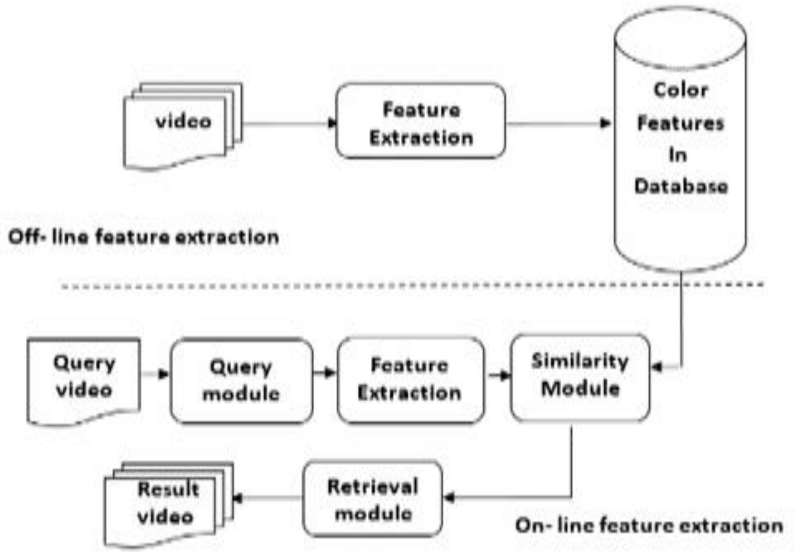

Figure 1. CBVR system modules

2.2. DCT feature extraction

(DCT) means discrete cosine transform. It has found applications for digital signal and image processing, and particularly for data compression/decompression transformation coding systems. In this work DCT used as shown in this flowchart Figure 2. The DCT process involves reading the video and then extract 10 frames from video and resize them to 256×256 pixels. After that the DCT equation is applied to the resized frame. (192) DC coefficients for each image are obtained and then stored in the system database.

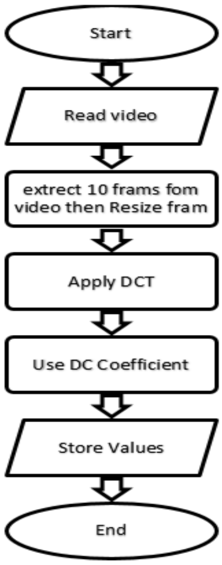

Figure 2. DCT feature extraction process

2.2.1. Frame extraction and resizing

Extracting 10 frames from video, according to the number of frames the required frames will be selected, resizing the input frame to fixed size, so all frame will be in new fixed width and high. Of all frames, the new size is 256×256. It is significant how this process gets the best result.

2.2.2. Discrete cosine transformation (DCT)

Applying DCT to the three frame matrices R, G and B to convert frame data from spatial domain to frequency domain by splitting each frame into 32×32 pixel blocks and apply DCT transform to each block value to produce 1024 coefficients. The output matrix consists of DCT coefficients, the valuable and important data representing the frame placed in the top left hand corner of the matrix while the less valuable details about the coefficients lies on the lower right-hand. In the top left hand corner of the matrix is DCT coefficient at position (0,0) which indicates the average of the other 1024 value in the matrix. The point (0,0) is called DC coefficient and the other point called AC coefficients. So we use the DC coefficient from each block for every frame in database and use it as DCT color feature.

2.3. Test material

The database of videos used in this paper belongs to our videos that collected from real world. video database consists of 8 classes each contains about 10 videos. So, the video tested total number is 100. These Video are of different sizes with color resolution 24 bit/pixel. Eight video categories of database are used (**Game**, **Person**, **Building**, **Children**, **Cats**, **Flowers, Beaches, Landscape**). Figure 3 shows Sample of video for collected Database.

Figure 3. Sample of video for collected database

3. RESULTS AND DISCUSSION

The performance of a retrieval system can be evaluated with respect to its precision and recall. Precision measures can be described as the system's ability to retrieve only relevant models; Whereas recall measures are represented the system's ability to retrieve all relevant models defined the number of relevant videos retrieved. In this paper, we evaluate 5 videos in each database class and for each category 3 videos retrieved will be similar as shown in the Table 1.

$$\text{Precision} = \frac{\textit{Number ofrelevent video retrived}}{\textit{Total number of video retrived}} = \frac{A}{A+B}$$

$$\text{Recall} = \frac{\textit{Number of relvent video retrived}}{\textit{Total number of rlevent video in database}} = \frac{A}{A+C}$$

Where A being the number of relevant videos retrieved, B being the number of irrelevant items and C being the number of relevant items not retrieved in the database. Figure 4 shows Sample of video retrieval results.

Table 1. The precision result

Class	Game	Person	Building	Children	Cats	Flowers	Beaches	Landscape	Average
Average Precision	0.4	0.86	0.6	0.53	0.66	0.8	0.6	0.73	0.6475

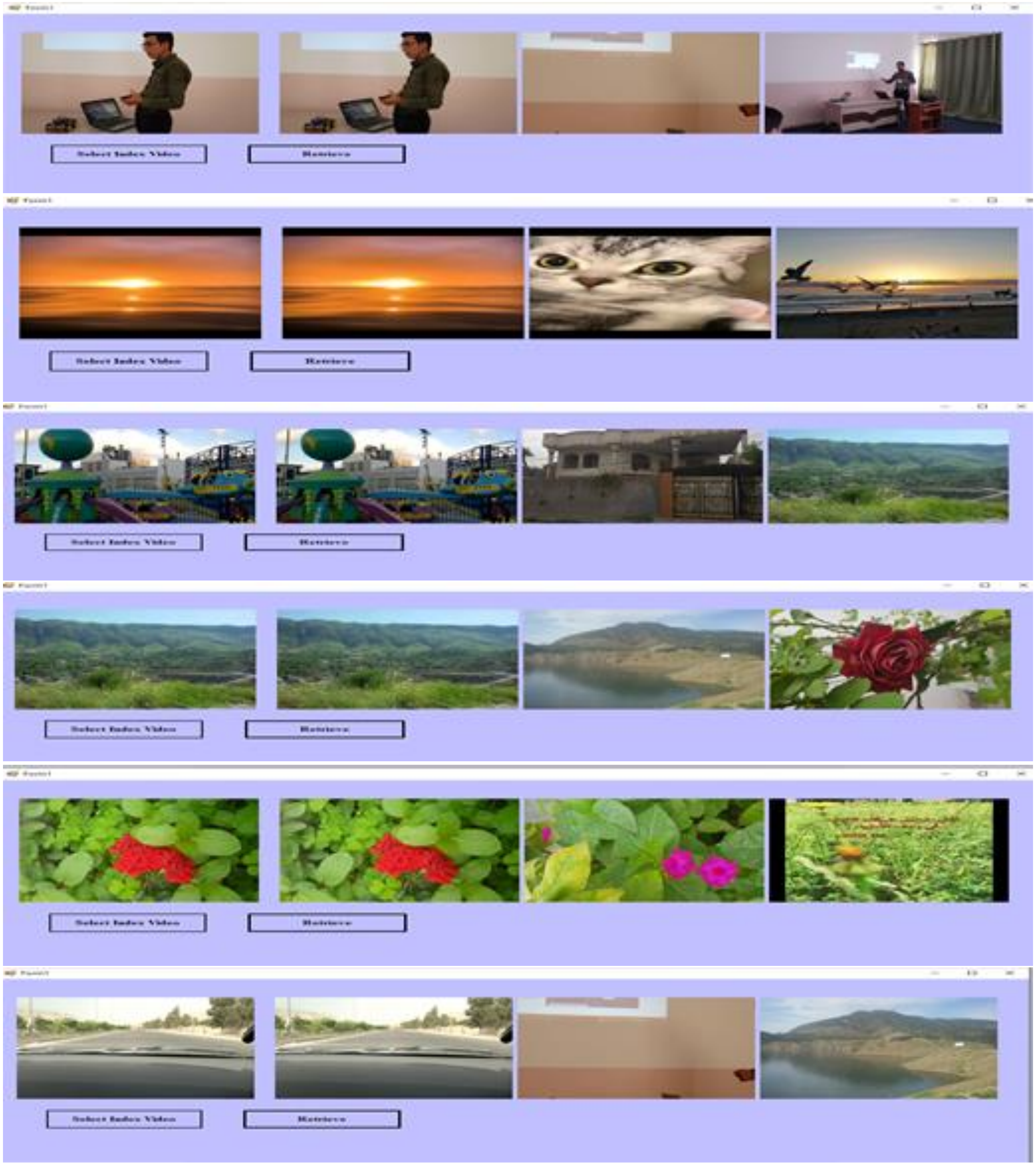

Figure 4. Sample of video retrieval results

4. CONCLUSION

The content-based model allows the system to acquire a video from a list of visual content-based videos that are entirely automatically extracted, such as colour, texture, shapes, but not attributes that are unrelated to the content. The visual content of this video carries little to no semantic content. In this paper, the color content of video is used to retrieve videos. The experimental results show that the average rating of 0.6475 was calculated using the DCT after applying it to our database for all categories, when it's used in content-based video retrieval (CBVR). As the implemented algorithms run on the video's color features and values, the division of video into blocks or pieces therefore gave more ability to retrieve the relevant videos.

REFERENCES

[1] S. T. M. Toriah, A. Z. Ghalwash, and A. A. A. Youssif, "Semantic-Based Video Retrieval Survey," *J. Comput. Commun.*, vol. 6, no. 8, pp. 28-44, 2018.
[2] S. S. Gornale, A. K. Babaleshwar, and P. L. Yannawar, "Analysis and Detection of Content based Video Retrieval," *Int. J. Image, Graph. Signal Process.*, vol. 11, no. 3, p. 43, 2019.
[3] R. Ashraf, M. Ahmed, U. Ahmad, M. A. Habib, S. Jabbar, and K. Naseer, "MDCBIR-MF: multimedia data for content-based image retrieval by using multiple features," *Multimed. Tools Appl.*, vol. 79, no. 13, pp. 8553-8579, 2020.
[4] L. Tong, R. Tong, and L. Chen, "Efficient retrieval algorithm for multimedia image information," *Multimed. Tools Appl.*, vol. 79, no. 13, pp. 9469-9487, 2020.
[5] R. Gasser, L. Rossetto, S. Heller, and H. Schuldt, "Cottontail DB: An Open Source Database System for Multimedia Retrieval and Analysis," in *Proceedings of the 28th ACM International Conference on Multimedia*, pp. 4465-4468, 2020.
[6] M. E. Abdulmunem and E. Hato, "Semantic Based Video Retrieval System: Survey," *Iraqi J. Sci.*, pp. 739-753, 2018.
[7] S. Putheti, M. N. S. Harsha, and A. Vishnuvardhan, "Motion Detection in Video Retrieval Using Content-Based Video Retrieval," in *Innovations in Computer Science and Engineering*, Springer, pp. 235-242, 2019.
[8] A. F. S. Devaraj *et al.*, "An Efficient Framework for Secure Image Archival and Retrieval System Using Multiple Secret Share Creation Scheme," *IEEE Access*, vol. 8, pp. 144310-144320, 2020.
[9] P. Saxena, S. K. Singh, and M. Srivastava, "Content-Based Retrieval of Multimedia Information Using Multiple Similarity Indexes," in *Soft Computing: Theories and Applications*, Springer, pp. 1235-1242, 2020.
[10] S. Heller, L. Sauter, H. Schuldt, and L. Rossetto, "Multi-Stage Queries and Temporal Scoring in Vitrivr," in *2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, pp. 1-5, 2020.
[11] H. Wang, Z. Li, Y. Li, B. B. Gupta, and C. Choi, "Visual saliency guided complex image retrieval," *Pattern Recognit. Lett.*, vol. 130, pp. 64-72, 2020.
[12] J. K. B. B.Ramu, "A Survey of Object Video Retrieval Systems Using Scale-Invariant Feature Transform (Sift)," *Int. J. Adv. Eng. Res. Dev.*, vol. 5, no. 3, pp. 1211-1216, 2018.
[13] N. Gayathri and K. Mahesh, "A Systematic Study on Video Indexing," *Int. J. Pure Appl. Math.*, vol. 118, no. 8, pp. 425-428, 2018.
[14] A. Ansari and M. H. Mohammed, "Content based video retrieval systems-methods, techniques, trends and challenges," *Int. J. Comput. Appl.*, vol. 112, no. 7, 2015.
[15] P. Chivadshetti, K. Sadafale, and K. Thakare, "Content based video retrieval using integrated feature extraction and personalization of results," in *2015 International Conference on Information Processing (ICIP)*, pp. 170-175, 2015.
[16] G. Cámara-Chávez, F. Precioso, M. Cord, S. Phillip-Foliguet, and A. de A. Araújo, "An interactive video content-based retrieval system," in *2008 15th International Conference on Systems, Signals and Image Processing*, pp. 133-136, 2008.
[17] J.-H. Su, Y.-T. Huang, H.-H. Yeh, and V. S. Tseng, "Effective content-based video retrieval using pattern-indexing and matching techniques," *Expert Syst. Appl.*, vol. 37, no. 7, pp. 5068-5085, 2010.
[18] D. Asha, M. Lata, and V. S. K. Reddy, "Content based video retrieval system using Multiple Features," *Int. J. pure Appl. Math.*, vol. 118, pp. 287-294, 2018.
[19] E. M. Saoudi, A. A. El Ouadrhiri, O. El Warrak, S. J. Andaloussi, and A. Sekkaki, "Improving Content Based Video Retrieval Performance by Using Hadoop-MapReduce Model," in *2018 23rd Conference of Open Innovations Association (FRUCT)*, pp. 1-6, 2018.
[20] M. Mühling *et al.*, "Content-based video retrieval in historical collections of the german broadcasting archive," *Int. J. Digit. Libr.*, vol. 20, no. 2, pp. 167-183, 2019.
[21] R. Bansal and S. Chakraborty, "Visual content based video retrieval on natural language queries," in *Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing*, pp. 212-219, 2019.
[22] M. A. Hassan, S. Saleem, M. Z. Khan, and M. U. G. Khan, "Story based video retrieval using deep visual and textual information," in *2019 2nd International Conference on Communication, Computing and Digital systems (C-CODE)*, pp. 166-171, 2019.
[23] L. Zhu, X. Lu, Z. Cheng, J. Li, and H. Zhang, "Flexible multi-modal hashing for scalable multimedia retrieval," *ACM Trans. Intell. Syst. Technol.*, vol. 11, no. 2, pp. 1-20, 2020.
[24] M. Otani, Y. Nakashima, E. Rahtu, and J. Heikkilä, "Uncovering Hidden Challenges in Query-Based Video Moment Retrieval," *arXiv Prepr. arXiv2009.00325*, 2020.
[25] N. Spolaôr, H. D. Lee, W. S. R. Takaki, L. A. Ensina, C. S. R. Coy, and F. C. Wu, "A systematic review on content-based video retrieval," *Eng. Appl. Artif. Intell.*, vol. 90, p. 103557, 2020.

BIOGRAPHIES OF AUTHORS

Sumaya Hamad received the B.Sc. (good) (first class) degree in computer science from University of Anbar, Anbar, Iraq, in 2002, and the M.Sc. degree in computer science from University of Anbar, Anbar, Iraq, in 2012. She is currently PhD student in Computer Science Department at University of Technology, Baghadad, Iraq, and she an Assist. teacher. at the College of Computer Science and Information Technology, University of Anbar. She has published 5 refereed journal and conference papers. Her current research interests include mobile computing, artificial intelligence,Ad Hoc networks, search engines, and information technology.

Ahmeed Suliman Farhan has received B.Sc. in Computer Science, Al-Anbar University, Iraq, (2007-2011). M.Sc degree (2012- 2014) in Computer Science Department, Al -Anabar University. Fields of interest: Multimedia. He taught many subjects such as cryptography, operation system, Robot System and networks.

Doaa Yaseen Khudhur received the B.Sc. (good) (first class) degree in computer science from University of Anbar, Anbar, IRAQ, in 2004, and the M.Sc. degree in computer science from University of Anbar, Anbar, IRAQ, in 2012. She is currently a teacher. at the College of Computer, University of Anbar. She has published 3 refereed journals. Her current research interests include E-commerce, cryptography, Data security, and information technology.