I.J. Wireless and Microwave Technologies, **2020, 4, 39-48**
Published Online August 2020 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijwmt.2020.04.05

A Deep Analysis of Image Based Video Searching Techniques

Sadia Anayat[1], Arfa Sikandar[2], Sheeza Abdul Rasheed[3] and Saher butt[4]
[1]GC Woman University Sialkot, Pakistan
[2]GC Woman University Sialkot, Pakistan
[3]GC Woman University Sialkot, Pakistan
[4]GC Woman University Sialkot, Pakistan
Email: cssadiaanayat@gmail.com

Received: 25 March 2020; Accepted: 02 May 2020; Published: 08 August 2020

Abstract: For many applications like brand monitoring, it's important to search a video from large database using image as query[1]. Numerous visual search technologies have emerged with the passage of time such as image to video retrieval(I2V), video to video retrieval(V2V), color base video retrieval and image to image retrieval. Video searching in large libraries has become a new area of research. Because of advance in technology, there is a need of introducing the well established searching techniques for image base video retrieval task. The main purpose of this study is to find out the best image based video retieving technquie. This research shows the importance of image base video retrieving in the searching field and addresses the problem of selecting the most accurate I2V retrieval technique. A comparison of different searching techniques is presented with respect to some characteristics to analyze and furnish a decision regarding the best among them. The accuracy and retrieval time of different techniques is different. This research shows that there are a number of visual search techniques, all those techniques perform same function in different way with different accuracy and speed. This study shows that CNN is best as compare to others techniques. In future, the best among these techniques can be implemented to reduce the searching time and produce the promising result.

Index Terms: I2V, V2V, Visual Search, Accuracy.

1 Introduction

Inflammable growth of visual data both online and offline and the outstanding success in web searching, expectation for image and video technologies has been increasing [23]. With advances in technology, there is a speedily growing amount of video related data collected in many applications. An new area of research is video searching in large libraries[25]. A huge amount of visual data is record and stored in the system of visual information. For image and video indexing productive and efficient techniques are needed. In visual information system almost the visual materials are keep in compressed form. It is prudent that we have to explore the different features extraction techniques. The image features extractions are execute on compressed images and videos without any decoding methods. Powerful tools are required for image query and features based image to go with keyword based technique for searching. Different type of image features such as color, object bases features and shape are extracted and it is stored as information[24].Multiple visual searches have come out with the passage of time such as video to video search, image to video search, object based video search, color based video search, texture based video search. But image and video searches face many challenges with the certain focus indexing and detection of large scale semantic concept[23]. In the area of image to video retrieval is one of the common problem in the area of video retrieval. By the solution of traditional image retrieval task many researches were inspired. For the image query task that automatically create object representation and return the object of interest in the form of video clips Sivic et al present a method for this in (2006). For the searching of related video segments different researches proposed different technique based on Bag of words framework [28]. Video retrieval is an area of research which focus with very little organized concern for users on technological aspects[29]. In most of the state of the art system for video retrieval, firstly we segment the video into shots and then we create single or multiple frames for every shot. Then an image is given as a query and the query image is matched with each shots. Conventional system is designed to find the videos that are exactly match with query image[30]. A video record includes a large amount of data and video shots. The most productive contribution in digital videos records is to provide well organized data so that the user can solve the problems visually. Image base video retrieval models help a lot to find the most promising result. Intuitive visualization interfaces have found censorious for successful video shots in recent TRECVID evaluation[23]. A video object is

described as a collection of video regions that are grouped together over multiple frames under some standard. The interesting and special contribution include in developing a complete automated algorithm of video analysis for different segmentation of object and features extraction[31].

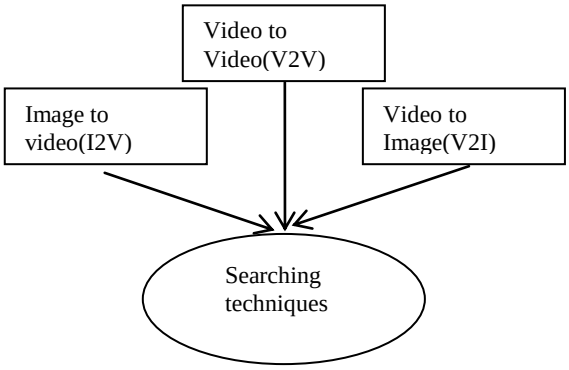

Fig. 1. Types of Visual Searching techniques

Visual search app have recently gained significant prominence[1]. As internet technology progresess, the number of video related data collected in many applications is growing rapidly. Image/video retrieval refer to the search for image/video that represent the same object as one shown in the image/video application[11]. Numerous visual search technologies have appeared in recent year such as video to image recovery(V2I), image to video(I2V) recovery and video to video(V2V) recovery[9].

Image to image(I2I) visual search can be used to search a product similar to the query image that is taken with the cell phone. Image to video(I2V) visual search is mostly used to retrieve most relevant video frame based on query image. Video to video(V2V) visual search is widely used in online video sharing websites to enforce copyright. This research discuss the image to video(I2V) retrieval technqiues such as CNN, VWII, etc. After a few year, a common user media library can contain thousand of videos. Such videos typically are not numbered, however, and eventually it become very difficult to scan specific content through them[10]. Usually an image/video is transmit by user to server to retrieve most similar image/video with that query[11]. There are multiple ways to perform this searching. The task may be completed through color base retrieval technique, content base retrieval techniques or feature base retrieval techniques. The color base searching is affected a lot by light. So, result may not be as promising as expecting. This study has been conducted to compare the techniques which are used for feature base image to video searching. We have studied different image base video searching techniques and compared them with each other with respect to some characteristics to find out which technique take less time and provide result with more accuracy and many more. Firstly, the image is given to the system as

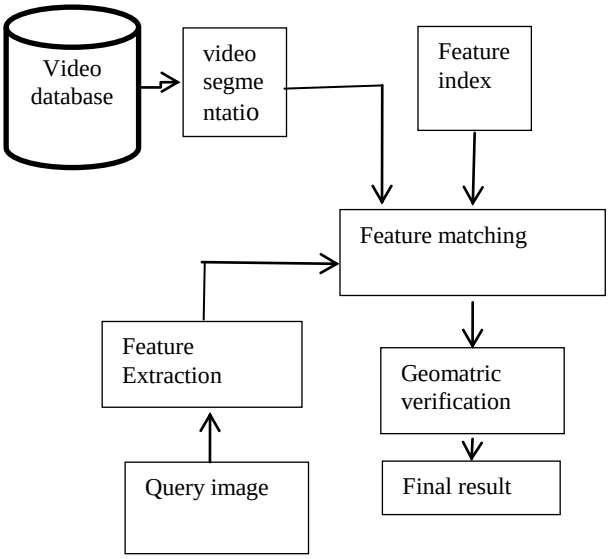

Fig. 2. A basic framework of image base video retrieval techniques(I2V)

query. Then, the specific features are extracted from query, Then matching models are used to match the image with the multiple videos in databse and the most relevant video is extracted. As compare to content base or color base searching, feature base searching produce most promising result. In this paper we have studied multiple image to video(I2V) retrieval models which perform searching by extracting the feature. The goal of this study is to examine the searching time and

strategy of different the image base video clip searching technique.

The above figure has shown the steps of video clips retrieval by feature extraction using image as query. Here, the image and video are used as query. The video is divided into multiple frames or segments and features are extracted from image. Then matching of features with video frames will be started and the most relevant frame or segment will be shown as output. Different techniques are used for this purpose. The basic goal of this research is to study those techniques and to find out the ebst among them.

This research is structured as follows: firstly, a brief introduction is provied in section 1. The related work is presented in section 2. Furthermore, analysis is provided in section 3. The discussion takes place in section 4 and draws conclusion at the end the paper in section 5.

2 Related Work

The problem of searching a large video database by image has been addressed in paper[1]. The basic goal of this study was to improve the retrieval scalability. The proposed archtecture have the ability to compare the query image directly with database videos. Retrieval scalability of proposed model was better as compare to baseline systems that search video frame level database. Then a video descriptor was introduced that could be compared with image descriptor directly. The proposed technique was improved up to 25%mAP for two kind of asymmetry [1].

The model was proposed for image base video retrieval task where CNN model was used for feature extraction of query image. The research explored the importance of using Region base convolutional neural network as a feature extractor. An image base video retrieval system that consists of filtering and re-ranking was introduced using proposal of object learned by RPN and its related information taken from CNN. The features on an object detected by CNN were used for image base video retrieval. The faster region base convolutional neural network was used as local and global descriptor. The finding of the proposed model was very positives[2].

In this paper, Different strategies have been discussed for aggregating video temporal information from video data based on the normal local features and profound representation of learning that focus on image to video recovery task. Aggregating local binary characteristics and deep learning functions are main focus of proposed model. The experimental result show that binary feature based approaches are better than deep learning approaches[3]

A semi--supervised ranking scheme which base on the graph was introduced for large scale video retrieval where multimedia tools were incorporated for better rating. The proposed model was enable to learn efficiently from small samples. For video retrieving from large database, a multimedia ranking framework was introduced where different ranking schmes are unifies. The experimental results showed that in TRECVID evaluation, the propsed scheme solution are efficient than state-of-art solution[4].

The problem of large scale application where indexing video frame of database independently is illogical have been focused in this paper. A bloom filter based system, for I2V retrieval task, was intoduced that can index long video segments. The proposed system was much efficient for image base video retrieval task. The experimental results showed that the proposed mode was efficient up to 24%mAP on public database and 4% faster than state-of-art solution[5].

A PHDL approach was proposed for image to video person re-identification problem. The proposed PHDL model can learn from training image-video pairs a FMP and a couple of heterogeneous dictionaries. The featured projection matrix used in the proposed technique could reduce variation within the videos[6].

Two heterogeneous spaces were embedded into a common discriminate hamming space and proposed a new novel heterogeneous hash learning model called HER which was used for face video retrieval using image as query. Discrimination, consistency and compatibility were explored during learning of hash function to iteratively refine cross space hash code[7].

The speeded up Robust Feature was used to introduced a model called Content base video retrieval. The proposed solution extract the similar frame from video according to query image. Recall, precision and running time were used to measure the performance of proposed model. Two measurements have been used to measure the performance of CBVR: Accuracy and Runtime. The system test within frame and not-in frame images. In Not-in frame, average precision was 25%, average recall was 66% and running time performance data gave 73.56 second. For in-frame, precision value was 59%, recall value was 51% and running time was 121.67 second in future, research will be on understanding the performance of SURF which is based on some specifics kind of videos[8]

A technique was implemented to solve top-k video retrieval problem through a query image. The basic purpose of this model is to retrieve best matched video from large video database using image as query. A model base on CNN and BOVW technique was proposed to improve the accuracy of video retrieval[9].

For video/image frames, a model called single fisher vector was proposed. The proposed technqiue was used for the recovery of best matched videos/image frames. An algorithm known as lossless was presented in which the number of arithmetic operation with the leanness of fishing vector decreased[10].

In this paper, to research effectiveness in video analysis, a novel deep learning feature has been established and intergrated them into well-established CDVA evaluation system. The performance of proposed model for video trieval, was effective. To obtain compact and strong CNN descriptor, a nested invariance pooling has been proposed in this paper.

The performance of proposed descriptor of CNN is better than all state-of-art CNN descriptor with gain of 11.3% mAP[11].

This paper has the vision and personal understanding of the authors. Audio visual information, enabling structural are encompass in MPEG-7 standard, ISO/IEC is the next standard inferior expansion by MPEG. Confession of multimedia content is established on MPEG-7, it comprise non MPEG and non-restrict arrangement. In September 2001, MPEG-7 desire to enhancing the international standard. MPEG-7 is used same expansion process that is used in MPEG. The component of MPEG-7 are: DDL,D, DSs. The stored media help the user to determine, retrieve, and filter audio visual information. Syntax and semantics present of audio visual content that is established on Ds, Ds is also include the color, texture, shape. DSs and Ds are occupying on XML scheme, DDL scheme is copied by expansion of XML scheme that are refined byW3C using XML. The MPEG-7 are access the storage on MP4files. SMPTE/EBU established on a dictionary model that consist about 1000 bit. The authors acknowledge the relevant comment from innominate analyst and A.Belen Benitez, which is the co-editor of the MPEG-7[27].

The JACOB system is determined in this paper. The proposed system is used to slash the video into the string of shots, abstract a slight frame from all shots and count these frames that occupying features like texture base and color base. The videos is stored in the required database, put a query the database shows a slight amount of video that is matched to the query. The system is divided into two functions population database and querying database. Using object descriptor better result absolutely achieves. The CHABOT system is used for storage and a lot of enormous retrieval of images. The content based storage and retrieve of images based on CANDID system[26].

Mathematical models that are video based are determined in this paper. The model is used for the video editing effect. They can be assigned by the feature extracting that acknowledges the cut, chromatic edit and spatial edit. A chromatic edit is accomplish by wield the intensity or color. Spatial edit is accomplished by conversion of the pixel space of shot that is edited. The digital video segmentation is present in this paper. The problem is that break the automation of camera location in string of videos. Video production techniques admit the production process that is used to the video segmentation design. The result is achieved using designed and classification approach. In the future the research will be inscribe to improve the problem of segmentation that is based on cut detection[25].

A sketch-based video recoup engine supporting multiple query model. The similarity measure of video contect is suggested in [17]. IMOTION system make use of a low level image and videos features and also make use of high level temporal and spatial features and they all are used in any type of mixture. For query statement the user have to provide the sketches of video in a frame and user also state motion across different frames and also provide the complete support for important feedback. This technique provides a very productive support for items which are already known in a large collection of videos[17].

A algorithm is presented for content based video retrieval which use edge and motion features. The algorithm using the K-mean method which extracts key frames from video that is followed by the extraction of edges and motion features that represent the feature vector and video. The proposed framework is compared with VLBP method. This method outplay well if we compare it with VLBL method. The proposed model VLBL is reasonably provide good result.[18]

The person re-identification plays an important role in observation and argumentative application of videos. A technique was proposed for identification of person from image to video .We see that there is exist a large variation in each video frames. These elements make matching very complex between image and video. They proposed joint feature PHDL approach for IVPR. The hold of variations with each video that is going to be match can be reduced with the use of learned projection matrix. This technique improve the effectiveness from image to video person identification[19].

Various types of audio and optical information that a video contain which are very difficult to merge and extract in video information retrieval. This technique provides the evaluation on various types information that are used for video recuperation from the collection of multiple videos. They discuss the benefaction of automatic speech recognition, matching the similarity in image, detection of face and video OCR in the condition of experiments. Image retrieval provides the best result for any type of single metadata[20].

Increasing of online videos the NDVR was gain many researcher's attention. There are many application of NDVR, like monitoring of online videos and copyright protection. The NDVR algorithm uses the representation of bag of features and exploits the local volume information for image base video retrieving. Such type of representation about global distribution are ignores the potentially precious information. Our reason is that when we use the global features for the classification of similar key frames which are very useful to improve the performance of NDVR. We proposed a technique which that is used to improve the random transform of features which provide the detail global distribution of interest point. It is calculated by two-dimension discrete RT. This approach outperform the state of the art in near duplicate videos[21].

The deep hashing framework technique namely UDVH, is used for retrieving the videos which are similar with each other with sight to learn dense further for productive binary codes. They produce the hash code by adding the integrating particular videos presentation with optimal code learning. UDVH generates the hash code by using the feature clustering with the original structure which is maintained in binary space. Large scale experiments performed on popular videos dataset which manifest that UDVH is best than the state of art in period of estimation metrics[22].

Below is the comparison of some Image base video retrieving technqies on the basis of their common characteristics.

Table 1. Comparison between image base video searching techniques

Reference	Techniques	Database used	Classifier	Results	Advantages	Disadvantages
[1]	Bloom Filter Frame work	S12V,VB, ClassX	Hashing and score computation,	Enable efficient and effective video retrieval,	achieve high retrieval accuracy and reduce query latency	Obtain limited retrieval accuracy
[2]	RPN,CNN	FERET [23], FACES94 [24]	RCN-Classifier	found that applying the similarity measure on the CNN feature gave the best results	reducing the feature extraction time	Take longer time in feature extraction
[3]	Temporal Aggregation of Hand-Crafted Features, Deep Learning Architecture	Movies DB	Temporal Aggregation Models of Visual Features, Feature Detection algorithm	Remove sacking between related images	Lump shot into single vector and take less memory	Not properly encrypt data for video retrieval by image query
[4]	Multi-Level ,Novel Multi-Modal technique	TRECVID benchmark	semi-supervised ranking (SSR) scheme	Provide the best retrieval performanc-e useful for large number of videos	The proposed work is fairly efficient for application of large scale	Not efficiently use in multimedia feild
[5]	*A. Asymmetric comparison techniques, Temporal aggregation*	SI2V-14M, VB-14M, ClassX-1.5M	Bloom Filter, Fisher vector	Enables Faster and more Memory-Efficient Retrieval	demonstrate up to 25% mAP improvement	Join high dimentional space
[6]	IVPR	iLIDS-VID, PRID 2011	Heterogeneous dictionary pair learning	Important in video surveillance	Feature projection matrix used to reduce variation in video	Point-to-set matchging problem
[7]	Hashing over Eulidean space and (HER)	Buffy the Vampire Slayer ,Big Bang Theory	Hashing over Eulidean space and HER are Two diverse spaces	First strive to sink two heterogene-ous space into a same hamming-space	Supposed model was impressively unique from state of art.	The experimental result was shown with 128 bit only because of space limitation
[8]	Content-base video retrieval by using features of speed up robust	descriptor database	Features of speed up robust database	Effort of system same for random categories for both videos and image.	.In Not-in frame, average precision was 25%, average recall was 66% and running time performance data gave 73.56 second. In Within frame, running was 121.67 second	the performance of SURF which is based on some specifics kind of videos is still need to improve. The samples are limited to contain specific predefined 5 categories.
[9]	Visual Weighted Inverted Index	V Database	Convolutional Neural Network, Bag of words	our approach outperforms the state of-the-art method	Improve the Efficiency and Accuracy of Retrieval Process	contras the level of video clip one by one is unsatisfactory
[10]	VRFP	Database of video a-priori	CNN	Fisher vector based on CNN measured similarity between web image and video frames effectively.	System use strategies that do not require any discriminative training for retrieving video.	Text is used to input query.
[11]	Nested variance pooling(NVP)	Image -net classification datasets	Deep learning features,CDVA	The proposed model reduced dimensional of CNN features for compactness of video descriptor	significance mAP gains of 11.3% and 4.7% respectively.	Deep learning technique can explore more in the video descriptor standard development

[26]	JACOB system	MPEG or QuickTime, feature DB	CHABOT system, CANDID system	Using this technique the result are very adequate and examining	Content based storage and retrieve of images based on CANDID system.	The automatic and booming segmentation antiquated proved that a very difficult task on static images
[25]	Mathematical models	Non	cut, chromatic edit and spatial edit	The result is achieve using designed and classification approach. The research will be inscribe to improve the problem of segmentation that is based on cut detection.	A video production techniques admit the production process that is used to the video segmentation design	The problem is that break the automation of camera location in string of videos.
[27]	MPEG-7	MP4files	XML Scheme, SMPTE/EBU	The result is that The stored media helps the user to determine, retrieve, filter audio visual information	SMPTE/EBU established on a dictionary model that consist about 1000 bit	The behavior of the confession exhausting system will not be regularizing and confide on the distinct application
[17]	low-level features, high-level spatial and temporal features	Image dataset	IMOTION system	Productive support for related feedback	productive support for items which are already known in a large collection of videos	Does not employ the learning method
[18]	K-mean method	dataset of 335 videos	CBVIR	reasonably good performance as compare to the VLBP	Perform well compare to the VLBP	Extract one frame at a time
[19]	IVPR,FPM	iLIDS-VID,PRID	PHDL	Variation in the video can be removed	We get better results with PHDL instead of state of the art	Only reduce the intra video variation
[21]	BOF	CC_WEB_VIDEO dataset	IR	Combine the both BOF with IR features	Outperform the state of the art	slow motion or picture in picture transformation cannot be retrieved accurately
[22]	UDVH	FCVID, YFCC, Activity Net	deep hashing framework	Improve accuracy and efficiency	Better than the state of art in term of estimation matrix	Manage the low dimension space

3 Analysis

11 best techniques using for image base video retrieval have chosen from above table on the basis of their speed and accuracy. All these technqiues give the best result for image base video searching. As VWII extract the features of given image and then match these features with the videframes to retrieve the best match, so the outcome of this model will be promised. In Asymmetric comparison techniques, speed is improved up to 25% mAP. Then CNN model reduce the time used for feature extraction. Multi-Level ,Novel Multi-Modal technique is fairly efficient for application of large scale. The Bloom filter Framework has also been choosed in 11 best technqiue because it reduces query latency and achieves high retrieval accuracy. In Content-base video retrieval by using features of speed up robust, the running time in within frame was 121.67 second, In Not-in frame, running time performance data gave 73.56 second. Speed is improved by 12.2 in IVPR for Image base video retrieval task. Stanford 12V is best because Feature projection matrix used to reduce variation in video. Temporal Aggregation of Hand-Crafted Features, Deep Learning Architecture model take less memory and chunks shot into a vector. So all these technqiues are compared in this study on the base of some important features like accuracy, challenges etc.

Table 2. Comparison between best 11 techniques among above

	Technique	**Accuracy**	**Classifier**	**Database used**	**Speed**	**Challenges**	**prons**
I	Visual Weighted Inverted Index	The Accuracy AND Efficienc of Retrieval Process is improved	CNN, BOW	V Database	3X speed-up faster in completing retrival task	Computationally expensive	Improve the Efficiency and Accuracy of Retrieval Process
B. I	Asymmetric comparison techniques, Temporal aggregation	25% mAP improvement	Bloom Filter, Fisher vector	SI2V-14M, VB-14M, ClassX-1.5 M	News videos 229 queries (from web) 2.7 minutes/clip 50.6 shots/clip	demonstrate up to 25% mAP improvement	Speed is improved up to 25% mAP
III	RPN,CNN	obtained Results are very positive	RCN-Classifier	FACES94 [24],FERET [23]	0.926 mAP	Take long time in feature extraction	reduce the time used for feature extraction.
IV	Multi-Level ,Novel Multi-Modal technique	0.8952 in mAP	semi-supervised ranking (SSR) scheme	TRECVID benchmark	Faster than state-of-the-art solutions in the TRECVID evaluations.	Not efficiently use in multimedia feil	The proposed work is fairly efficient for application of large scale
V	Bloom Filter Frame work	25% mAP improved	Hashing and score computation,	S12V,VB, ClassX	Speed is up to 4×, 9.6× and 5.6×, for the VB-4M, SI2V-4M and ClassX-1.5M datasets, respectively.	Obtain limited retrieval accuracy	reduce query latency and achieve high retrieval accuracy
VI	IVPR	Introduced IVPR problem	Heterogeneous dictionary pair learning	iLIDS-VID, PRID 2011	Speed is improved by 12.2% (=36.8%-24.6%)	Point-to-set matchging problem	Feature projection matrix used to reduce variation in video
VII	Hashing over Eulidean space and (HER)	low accuracy tradeoffs	Hashing over Eulidean space and Riemannian manifold (HER) Are Two diverse spaces	Big Bang Theory, Buffy the Vampire Slayer	300 is a trade-off between computational cost and retrieval accuracy.	because of space limitation, The experimental result was shown with 128 bit only	The proposed solution was impressively unique from state of art.
VIII	temporal aggregation of hand-crafted features(SIFT, BRIEF)	Edge ‘information is also incorporated into the segmentation process to improve the accuracy.	RNN, GRU, LSTM, CNN,	VB-600k, SI2V-600k, Movies DB	LSMDC dataset contains 118 short video clips and 202 movies split into 128 of about 5 seconds.	The process of Features extraction is slow	Efficient in the total cost and precision.
IX	Content-base video retrieval by using features of speed up robust	In Not-in frame, average precision was 25%, average recall was 66% and running time performance data gave 73.56 second. In Within frame, running was 121.67 second	Features of speed up robust database	descriptor database	running time in within frame was 121.67 second, In Not-in frame, running time performance data gave 73.56 second	the SURF performance (based on some specifics kind of videos) is still need to improve.	.In Not-in frame, average precision was 25%, average recall was 66% and running time performance data gave 73.56 second. In Within frame, running was 121.67 second

X	Stanford 12V	The accuracy of this model is high because 3, 800 hours of newscast videos and annotated more than 200 ground-truth queries are collected	temporal refinement stage, Scene retrieval stage	Stanford 12V	Feature projection matrix used to reduce variation in video	More storage space is required to datasets	baseline for future research for retrieval purpose is provided
XI	Temporal Aggregation of Hand-Crafted Features, Deep Learning Architecture	24 to 30 frames per second	Feature Detection algorithm, Temporal Aggregation Models of Visual Features	Movies DB	As compare to Euclidean distance for matching SIFT Vectors, the Hamming distance for matching binary features is faster to compute	SI2V-600k, VB-600k, Movies DB	take less memory and chunks shot into a vector

4 Discussion

Inflammable growth of visual data both online and offline and the outstanding success in web searching, expectation for image and video technologies has been increasing. Image to video retrieval is one of the common problems in the area of video retrieval. Multiple technqiues are being used for I2V. We have compared some techniques in table 1 on the basis of their common characteristics. The above comparison shows the diffrence between the technqiues used for video retrieval purpose. Each technique has unique characteristics. Image is being used as query in all these models the the retrieval time, accuracy and strategy for searching the best matched video is different. These techniques are use for same purpose but in different manners. The features that make these techniques diffrent from each other may be the speed, accuracy, etc. The comparison is presented to show the cons, prons and functionality of these models.

VWII is a technique that improves the accuracy and efficiency of retrieval process and 3x speed up but it is time undesirable to Contrast the level of video clip one by one. Asymmetric comparison techniques are fast up to 25%mAP. CNN and RCN techniques are best for reducing the time use for feature extraction. The results obtained from this model are very positive and retrieval speed is improved up to 0.9%mAP. Bloom Filter technique use hashing and secure computation and improve the speed of retrieval process up to 25%mAP. This model reduces query latency and achieves high retrieval accuracy. In IVPR model, the problem of image to video person re-identification was addressed. This model use Feature projection matrix to reduce variation in video. Hashing over Eulidean space and (HER) provide low accuracy trade off. Because of space limitation of model, the experimental result was shown with 128 bit only. But this model is impressively unique from state-of-art solution. The process of Features extraction is slow in temporal aggregation of hand-crafted features(SIFT,BRIEF). In this solution, the Edge ‘information is also incorporated into the segmentation process to improve the accuracy. But this model is efficient in low cost and percision. In Content-base video retrieval by using features of speed up robust, running time in within frame was 121.67 second, In Not-in frame, running time performance data gave 73.56 second. In Not-in frame, average precision was 25%, average recall was 66% and running time performance data gave 73.56 second. In Within frame, running was 121.67 second. In Stanford 12V, there is storage issue because more storage space is required to datasets but it is faster up to 20% by removing 2 code words. The accuracy of this model is high because 3, 800 hours of newscast videos and annotated more than 200 ground-truth queries are collected. Temporal Aggregation of Hand-Crafted Features, Deep Learning Architecture search 20 to 30 frames per second. This model takes less memory and chunks shot into a vector.

This study shows that features are extracted in CNN model. Then these features are compared with each frame of video and most relevant frame/segment is retrieved. The results of this technique will be more promising. CNN is also best for reducing the time use for feature extraction. The results obtained from this model are very positive and retrieval speed is improved up to 0.9%mAP. So, the CNN is the best from other technqiues because of having the characteristic of feature extraction.

5 Conclusion

This work addresses the problem of selecting the most accurate image base video retrieval technique. With advance in technology, there is a need of finding the most efficient technique for image and video searching. A lot of techniques have

been introduced by different people in different eras for video retrieval by image. This study has been conducted to flash light on different image base video retrieval techniques. Reviewing these retrieval models may reflect different light in searching field. This study has given the better picture of which technique is better for video retrieval task using image as query. The comparison of different searching techniques is presented in this research with respect to some common characteristics to find out which searching model produce most promising results in short time period. Then the 11 best techniques among them, on the basis of their characteristics, are compared again to find out the most effective technqiue for image base video retrieval purpose. This research shows that CNN model is best for video retrieval task as it reduce feature extraction time and increase the retrieval speed upto 0.9%mAP. This technique can be implemented to get the best matched video from the database by matching the features. In future, with the emergence of deep learning techniques and large video datasets, image base video retrieval will be promising research direction. The work can be expanding to retrieving the clips from video using image as query.

References

[1] Araujo, Andre, and Bernd Girod. "Large-scale video retrieval using image queries." *IEEE transactions on circuits and systems for video technology* 28.6 (2017): 1406-1420.

[2] Hachchane, Imane, et al. "Video retrieval with CNN features." 2019.

[3] Garcia, Noa. "Temporal aggregation of visual features for large-scale image-to-video retrieval." *Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval*. 2018.

[4] Hoi, Steven CH, and Michael R. Lyu. "A multimodal and multilevel ranking scheme for large-scale video retrieval." *IEEE transactions on Multimedia* 10.4 (2008): 607-619

[5] Araujo, Andr é, et al. "Large-scale query-by-image video retrieval using bloom filters." *arXiv preprint arXiv:1604.07939* (2016)

[6] Zhu, Xiaoke, et al. "Learning heterogeneous dictionary pair with feature projection matrix for pedestrian video retrieval via single query image." *Thirty-First AAAI Conference on Artificial Intelligence*. 2017

[7] Li, Yan, et al. "Face video retrieval with image query via hashing across euclidean space and riemannian manifold." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015.

[8] Tarigan, Jos Timanta, and Evi P. Marpaung. "Implementing Content Based Video Retrieval Using Speeded-Up Robust Features." International Journal of Simulation–Systems, Science & Technology 19.3 (2018)

[9] Zhang, Chengyuan, et al. "CNN-VWII: An efficient approach for large-scale video retrieval by image queries." *Pattern Recognition Letters* 123 (2019): 82-88.

[10] Han, Xintong, et al. "VRFP: On-the-fly video retrieval using web images and fast fisher vector products." *IEEE Transactions on Multimedia* 19.7 (2017): 1583-1595.

[11] Lou, Yihang, et al. "Compact deep invariant descriptors for video retrieval." *2017 Data Compression Conference (DCC)*. IEEE, 2017.

[12] Ho, Yu-Hsuan, et al. "Fast coarse-to-fine video retrieval using shot-level spatio-temporal statistics." *IEEE Transactions on Circuits and Systems for Video Technology* 16.5 (2006): 642-648.

[13] Ma, Wei Y., Yining Deng, and B. S. Manjunath. "Tools for texture-and color-based search of images." *Human Vision and Electronic Imaging II*. Vol. 3016. International Society for Optics and Photonics, 1997.

[14] Smith, John R., and Shih-Fu Chang. "Image and video search engine for the world wide web." *Storage and Retrieval for Image and Video Databases V*. Vol. 3022. International Society for Optics and Photonics, 1997.

[15] Zhang, Ruofei, et al. "Video search engine using jointcategorization of video clips and queries based on multiple modalities." U.S. Patent Application No. 11/415,838.

[16] Wen, Che-Yen, Liang-Fan Chang, and Hung-Hsin Li. "Content based video retrieval with motion vectors and the RGB color model." *Forensic Science Journal* 6.2 (2007): 1-36.

[17] Rossetto, Luca, et al. "IMOTION—a content-based video retrieval engine." *International Conference on Multimedia Modeling*. Springer, Cham, 2015.

[18] Ravinder, M., and T. Venugopal. "Content-Based video indexing and retrieval using key frames texture, edge and motion features." *International Journal of Current Engineering and Technology* 6.2 (2016): 672-676.

[19] Zhu, Xiaoke, et al. "Learning heterogeneous dictionary pair with feature projection matrix for pedestrian video retrieval via single query image." *Thirty-First AAAI Conference on Artificial Intelligence*. 2017.

[20] Hauptmann, Alexander G., Rong Jin, and Tobun D. Ng. "Video retrieval using speech and image information." *Storage and Retrieval for Media Databases 2003*. Vol. 5021. International Society for Optics and Photonics, 2003.

[21] Liao, Kaiyang, et al. "IR feature embedded bof indexing method for near-duplicate video retrieval." *IEEE Transactions on Circuits and Systems for Video Technology* 29.12 (2018): 3743-3753.

[22] Wu, Gengshen, et al. "Unsupervised deep video hashing via balanced code for large-scale video retrieval." *IEEE Transactions on Image Processing* 28.4 (2018): 1993-2007

[23] Chang, Shih-Fu. "Compressed-domain techniques for image/video indexing and manipulation." *Proceedings., International Conference on Image Processing*. Vol. 1. IEEE, 1995.

[24] Chang, Shih-Fu, Wei-Ying Ma, and Arnold Smeulders. "Recent advances and challenges of semantic image/video search." *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07*. Vol. 4. IEEE, 2007.

[25] Hampapur, Arun, Terry Weymouth, and Ramesh Jain. "Digital video segmentation." *Proceedings of the second ACM international conference on Multimedia*. 1994.

[26] La Cascia, Marco, and Edoardo Ardizzone. "Jacob: Just a content-based query system for video databases." *1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings*. Vol. 2. IEEE, 1996.

[27] Chang, Shih-Fu, Thomas Sikora, and Atul Purl. "Overview of the MPEG-7 standard." *IEEE Transactions on circuits and systems for video technology* 11.6 (2001): 688-695.
[28] Zhang, Chengyuan, et al. "CNN-VWII: An efficient approach for large-scale video retrieval by image queries." *Pattern Recognition Letters* 123 (2019): 82-88.
[29] Smeaton, Alan F., et al. "Collaborative video searching on a tabletop." *Multimedia Systems* 12.4-5 (2007): 375-391.
[30] Foote, Jonathan T., Andreas Girgensohn, and Lynn Wilcox. "Methods and apparatuses for interactive similarity searching, retrieval, and browsing of video." U.S. Patent No. 6,774,917. 10 Aug. 2004.
[31] Chang, Shih-Fu, et al. "VideoQ: an automated content based video search system using visual cues." *Proceedings of the fifth ACM international conference on Multimedia*. 1997.

Authors' Profiles

Sadia Anayat was born in Sialkot in 1998. She did her matriculation level in 2014 from, Govt. Girls Higher Secondary school Sambrial, district Sialkot in Science Subjects and her intermediate level (I.C.S) in 2016 from Superior college Sambrial, district Sialkot. Now she is doing her BS (Hons) in Computer Science (CS) from GCWU Sialkot. She is also certified as Microsoft Office specialist. Currently she has been working on research in Blochain technology and comparison between bitcoin and ethereum topic. Her main areas of research interest are blockchian technology, image processing and WSN.

Sheeza Butt was born in Sialkot in 1999. She did her matriculation level in 2014 from, Govt. Girls Higher Secondary school Sambrial, district Sialkot in Science Subjects and her intermediate level (I.C.S) in 2016 from Superior college Sambrial, district Sialkot. Now she is doing her BS (Hons) in Computer Science (CS) from GCWU, Sialkot. She is also certified as Microsoft Office specialist. Currently she has been working on research in Blochain technology and comparison between bitcoin and ethereum topic. Her main areas of research interest are blockchian technology, image processing and WSN.

Saher Butt was born in Sialkot in 1999. She did her matriculation level in 2014 from Danish Public High School Sambrial, Sialkot in Science Subjects and her intermediate level (I.C.S) in 2016 from Superior College Sambrial, Sialkot. Now she is doing her BS (Hons) in Computer Science (CS) from GCWU, Sialkot. She is also certified as Microsoft Office specialist. Currently she has been working on research in Blochain technology and comparison between bitcoin and ethereum topic. Her main areas of research interest are blockchian technology, image processing and WSN.

Arfa Sikandar is a lecturer in Govt. College Women University Sialkot. Her main area of research are image processing and artificial intelligence.

How to cite this paper: Sadia Anayat, Arfa Sikandar, Sheeza Abdul Rasheed, Saher butt, " A Deep Analysis of Image Based Video Searching Techniques ", International Journal of Wireless and Microwave Technologies(IJWMT), Vol.10, No.4, pp. 39-48, 2020.DOI: 10.5815/ijwmt.2020.04.05