International Journal of Advances in Scientific Research and Engineering (ijasre)

E-ISSN : 2454-8006

DOI: 10.31695/IJASRE.2023.9.12.7

Volume 9, Issue 12
December - 2023

Advanced Evaluation Method for Video Retrieval System

Sally Ali Abdulateef

Computer Science Department

University of Al-Mustansiriyah

Baghdad, Iraq

ABSTRACT

In the past few years, the multimedia storage has increased and costs of storing multi-media data has become cheaper. Which is why, there are large numbers of videos that are available in the video repositories. With development of the multi-media data types and available bandwidths. The proposed method transfers each video of database into scenes using color histogram based scene change detection algorithm and key frames are extracted. For key frames multiple features are obtained using straight forward rules. A new framework depending on CNNs (convolutional neural networks) is recommended to perform the content based video retrieval with the less storage cost also with higher search capability. The recommended framework subsists of extraction algorithm with respect to key-frame and the feature collection strategies. Particularly, the extraction algorithm of key-frame takes benefit of clustering idea; so in that case excessive information is taken out from video data and also the storage cost is highly shortened. This work present a method uses the extracted features with convolutional neural network (CNN) for classification tasks. In this research paper, different types of videos will be used and the important features will be extracted using SIFT, after that we use the deep learning(CNN) process will be performed. The experimental results showed the efficiency and effectiveness of suggested approach.

Keyword: Convolutional Neural Networks (CNN), Feature Extraction (FE), Scale-Invariant Feature Transform (SIFT), Video retrieval (VR).

1. INTRODUCTION

In the previous years, recent trends in the video industry, the development of information technologies and the development of the multi-media strategies, the amounts of the accessible multi-media data are exponentially increasing. Video is a very important form of the multi-media data; [1,2]. Large amounts of the videos are acquired then stored on the computer devices because of the widespread uses of the digital video devices in a variety of the regions. The capability for the manipulation and access of the stored videos is commonly utilized by the users of various domains in a variety of the ways [3,4]. The systems of video retrieval have gained great importance as a result of an increasing amount of the visual data for the videos [5]. With no efficient and effective retrieval systems, it would not be possible to cope with increases of World Wide Web and also as a result of the advancement of the digital information [6,7]. There is more than one way to retrieve the video, including methods that depend on the semantic meaning and other methods that depend on the shape, size and texture [8]. Video information may be classified to 3 categories, which are: information of the low level features, representing visually and aurally, Syntactic information which describe the video contents, and Semantic information which describe what happens in the video based on the perception of the users. Semantic information that is utilized for the determination of the video events include: spatial information that is provided by a video frame, such as the location and objects that are provided in the video frame, in addition to temporal information that is provided by series of the video frames in timely manners, such as the movements and actions of an object that has been shown in the frame series [9-11]. Deep learning (DL) can be defined as a type of artificial intelligence (AI) and machine learning (ML) imitating how the human being gain specific knowledge types. DL is one of the important elements of the data science, including the predictive modeling and statistics [12]. It's highly advantageous for the data scientists that have been tasked with the collection, analysis and

interpretation of large data amounts; DL makes this procedure easier and faster [13]. The various NN types in the DL, like the CNNs, recurrent NNs (RNNs), ANNs, etc. have been changing how people interact with their world. As a result of the fast advancement in deep learning, more approaches that have been based upon DL provided us with prospects of the sufficient and precise retrieval of the videos. DL has the ability of acquiring semantic features of high-level through the combination of the visual features of lower-level and Deep CNNs had proven to be multipurpose tools for image representation with strong capabilities for the generalization, Rapid retrieval of the videos based upon deep NNs has been receiving more attention [14]. In the present study, different types of videos will be used and the important features will be extracted using SIFT, after that we use the deep learning(CNN) process will be performed.

In the following sections the proposed method will be presented in detail, section 2 view related works, section 3 Video Terminology, section 4 Feature Extraction section 5 explains proposed system, section 6 presents experimental results in the test, and section 7 views the conclusions.

2. RELATED WORK

Zhang et al. [15] proposed design a motion pattern descriptor to indicate features of the motion of the video in general way. Support Vector Machines (SVMs) are employed to assign motion texture to semantic concepts. The results show that motion texture is effective and compact in order to represent motion pattern as well as is improving motion based shot retrieval performance as a result of motion pattern descriptor's comprehensiveness and ability of semantic classification. Yang, et al. [16] proposed crowd video retrieval system with the use of the hand drawn sketches as queries. The difficulty in this work is representation of the crowd motion and measurement of similarity therefore; the algorithm of the motion structure coding is used for the motion level crowd video indexing and sketch representations and distance metric fusion technique that has been incorporated with Ranking SVM is utilized for the measurement of relevant degree between sketch query and the videos of the motion crowds. The experimental results show that the suggested approach has the robust and effective and outperforms of retrieval performance than alternative methods. Aharon et al. [17] proposed over complete LR and HR dictionaries have been trained jointly on the LR and HR image patches. Every one of the LR image patches may be depicted as scarce linear combination of the atoms from an LR dictionary. The dictionaries have been coupled through some common coefficients, which are also known as the representation weights. The coefficients as well as dictionaries may be found with the standard sparse coding approaches, like the K-SVD. Lu et al. [18] have carried out successive frames in video data as NN input data convolution, so that one has the ability of introducing data on time dimension, for the purpose of identifying the human body motions. Kim in [19], have sketched a number of the typical models of CNN that have been applied to the feature extraction in the classification of texts, and filter with a variety of the lengths, utilized for convolving the text matrix. The filter widths equal to the word vector lengths. After that, max pooling has been utilized for operating the extractive vectors of each one of the filters. Ultimately, every one of the filters is corresponding to some digit and connects those filters for the purpose of obtaining a vector that represents that sentence, which final prediction has been based on. Kalchbrener, et al [20], have produced a model which is rather complex, where the convolutional operation of every one of the layers is followed by process of max pooling.

3. VIDEO TERMINOLOGY

Prior to entering in discussion, it is sensible to initially highlight the videos' levels of hierarchy [21] as it has been depicted in Figure1.

- Video: which means a source of multi-media combining a stream of the images for the formation of animated picture, a audio component corresponding to the images which have been shown on screen and text data which is rendered with the linguistic forms.
- Key frame: because of similarity amongst the subsequent frames, one or several key frames have been chosen from one shot, on the basis of shot content complications. The chosen key frames denote salient visual content.
- Shot: which represents subsequent frames that have been captured with one camera with no considerable variations in the visual contents. It represents a video stream brick. Shot boundary detection represents a video track partition to the shots for enabling a variety of the operations of processing on the video.

DOI: 10.31695/IJASRE.2023.9.12.7

International Journal of Advances in Scientific Research and Engineering (ijasre)

E-ISSN : 2454-8006

DOI: 10.31695/IJASRE.2023.9.12.7

Volume 9, Issue 12
December - 2023

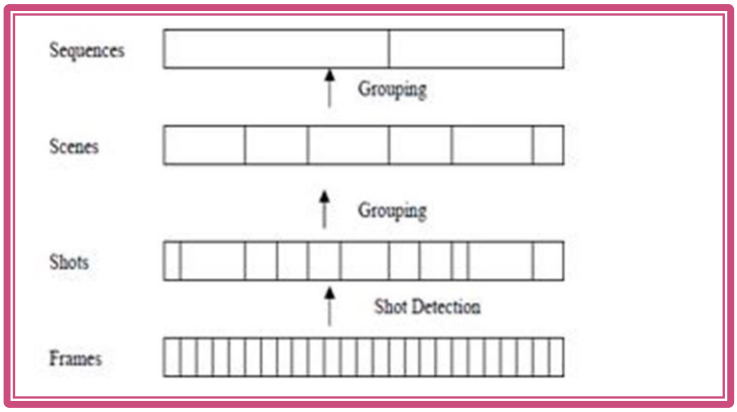


Fig. 1: General Hierarchy of Video Parsing

3.1. Basic Video Retrieval Concepts

- Frames: a video is divided to frames (i.e. images).
- Feature Extraction: the features are obtained from an image with the use of the sift approaches.
- Matching: In 3rd step those obtained features are matched from data-base videos.
-

4. IMPORTANT FEATURE EXTRACTION

It may interest many researchers to choose the appropriate features used in applications.

- Upload your video files.
- Specify the basic details of the task. The data to be collected and the keys used to carry out the work.
- You can add labels, lines, and shapes to make it clear which parts of the video need attention.

SIFT is utilized for the academic researches, it is a feature descriptor as well as a feature detector [22]. It can transform the image to a large set of the vectors of the local features, every one of which has invariance to the scaling, translation, and rotation of the image, as well as a partial invariance to the changes of illumination and affine or 3-D projection. Older methods to the generation of the local feature lacked the invariance to the scaling and have been with a higher sensitivity to the changes of the illumination and projective distortion [23,24,25]. SIFT features are sharing several characteristics in common with responses of the neurons in the inferior temporal (IT) cortex in the primate vision. The author of SIFT has also described enhanced methods to the indexing as well as model verification. SIFT features have been identified efficiently with the use of staged filtering method [26]. The 1st one of the stages identifies the key locations in the space of the scale by searching for the locations which are difference-of-Gaussian function's minima or maxima. Every one of the points has been utilized for the generation of feature vector describing local image area that has been sampled based on its scale-space coordinate frame. Features achieve a partial invariance to the local changes, like the affine or 3-D projections, through the blurring of the locations of the image gradient. [27, 28, 29,30]. Scale-Invariant Feature Transform (SIFT) is one of well-known image matching algorithms that work based on the local features in images .Key points for each object are extracted in order to provide "feature description" of the object [31]. SIFT contains the following steps [32].

- Constructing a Scale Space: To make sure that features are scale-independent
- Key point Localization: Identifying the suitable features or key points
- Orientation Assignment: Ensure the key points are rotation invariant
- Key point Descriptor: Assign a unique fingerprint to each key point

Finally, we can use these key points for feature matching!

The key points are identified using Difference of Gaussian "DoG" over different scales of the image. In other words, the image's scales and locations that differ the views of the same object are identified by searching the fixed features over all scales using scale space function. Different octaves from the original images are scaled to measure Gaussian value for each scale.

The scale space function is measured using Eq. (1) [32]:

$$L(x,y,\sigma)=G(x,y,\sigma)*I(x,y) \quad (1)$$

where * indicates the convolution between two scales (x and y). G (x, y, σ) is the scale Gaussian, which is measured using Eq. (2) [32]:

$$G(x,y,\sigma)=12\pi\sigma 2e-(X^2+Y^2)/2\sigma^2 \quad (2)$$

Finally, in order to get the scale-normalized Laplacian of Gaussian, the difference of Gaussian is measured using Eq. (3) [32]:

$$D(x,y,\sigma)=(G(x,y,k\sigma)-G(x,y,\sigma))*I(x,y)=L(x,y,k\sigma)-L(x,y,\sigma) \quad (3)$$

After measuring the Difference of Gaussian, the local extrema detection step is started to find the minima and maxima of Difference of Gaussian. The original image is blurred by Gaussian several times with changing the σ value in order to compare the X point in the original image with the above and below scales

Each point (x) is compared with (3 × 3) neighbors in the same scale, then it is also compared with (3 × 3) in the high and low scale. So, we have 26 pixels to be compared with (x). If (x) is larger or smaller than the 26 pixels, (x) will be considered as SIFT interesting point. The number of scales that must be taken is 3 scales; because they captured the most of interesting points in 3 scales as the experimental analysis proves [32]. Hence, the number of interesting points is normalized by applying a thresholding on minimum contrast and eliminating the outliers (i.e., edge responses) using Hessian matrix or Harris detector. Then, the orientation assignment is applied to achieve rotation invariance using two functions, Magnitude (Eq. (4)), and Direction (Eq. (5)) respectively [32].

$$m(x,y)=\sqrt{(L(x+1,y)-L(x-1,y))^2+(L(x,y+1)-L(x,y-1))^2} \quad (4)$$

$$\theta(x,y)=\tan^{-1}((L(x,y+1)-L(x,y-1))/(L(x+1,y)-L(x-1,y))) \quad (5)$$

where m is the magnitude, θ is the orientation between two scales (x, y). As shown in Fig. 2, the orientation of each interest point is identified according to the direction of the interest point peak in the histogram. In case there is more than one peak, multi direction will be found. But, unique direction (i.e, unique peak) is generally exist for gradient.

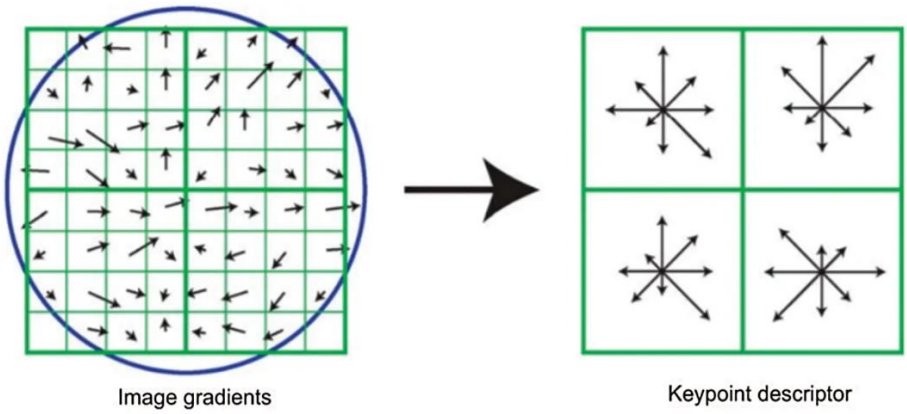


Fig.2 : Image gradients and key points descriptors [32]

DOI: 10.31695/IJASRE.2023.9.12.7

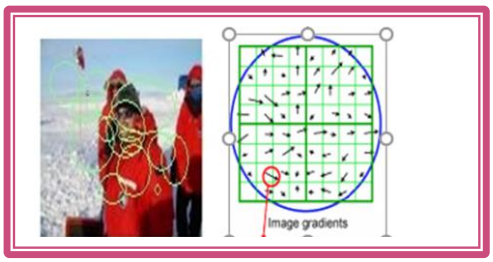


Fig .3: Example of SIFT Implementation

5. PROPOSED METHOD

A video database [40] that provides high-quality films will be used. They are used to train and test our algorithm. Feature extraction is the main core of diagnosis, classification, clustering, identification and detection. SIFT algorithm has been first suggested in 1999 by D. Lowe. Its applications include robotics mapping and navigations, object recognition, 3-D modeling, image stitching, video tracking and gesture recognitions. In the present research, SIFT feature was utilized, for the purpose of applying it to the video after the video is divided into frames. In the proposed framework, based upon the object from data-base utilizing the features of SIFT. The task of video classification is compiling the videos along with similar content and after that, assigning the videos to pre-defined class under supervision. In video genre classifications, the videos are categorized into different genres like news, movie, cartoon and sports.

Video genres classification is using widely previous knowledge as well as the using of low level features due to the robustness of these features for video diversity. Now, we used convolutional deep learning. A CNN represents feed-forward NN, which is utilized in general for analyzing the visual images through the processing of the data with the grid-like topologies. In the present paper, there are two steps, learning and testing. In the former one, the data-set will be learnt to lead to a model of classification utilized in the testing. The pre-processing for every one of the video feature extractions utilizing SIFT, vocabulary construction with the use of the K-means and then classification with the use of the CNNs. In the step of testing, it's similar to previous steps. This step's result is the class label for any available testing data.

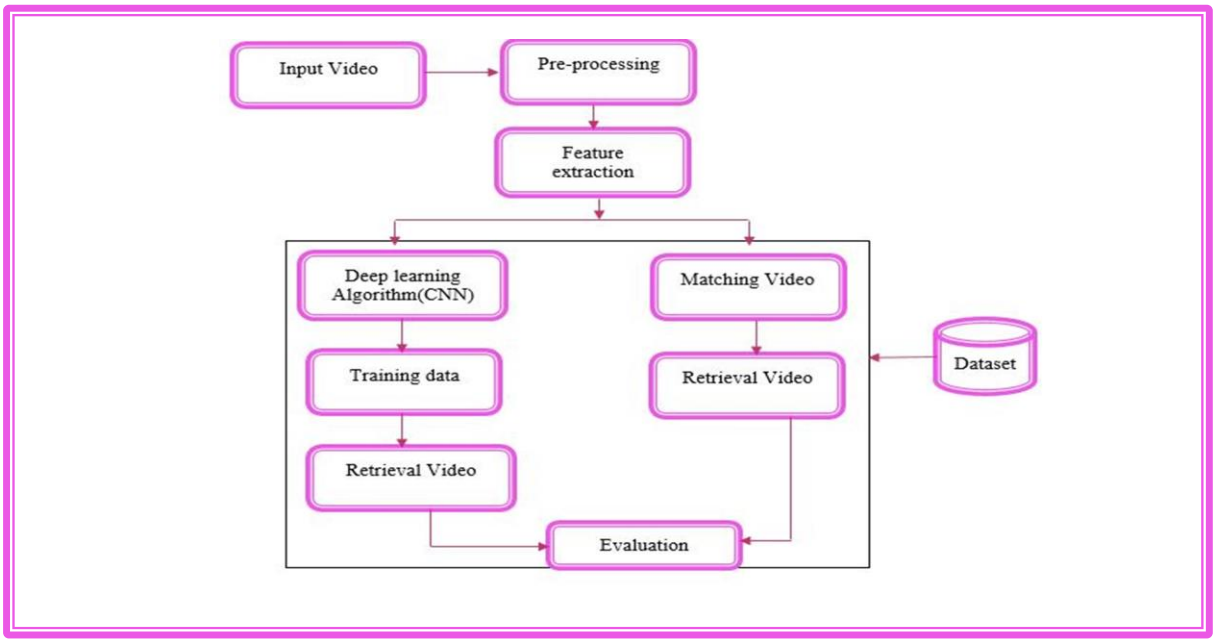


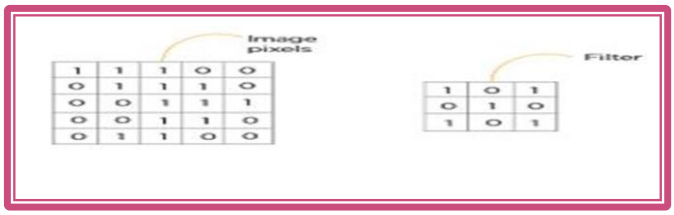
Fig.4: The Proposed Method Diagram

DOI: 10.31695/IJASRE.2023.9.12.7

5.1 Steps of New Algorithm

- Applying the convolutional filter in first layer for each frame.
- Obtain and Maintain the list of the last K predictions.
- The transfers of the signal from a layer to another is regulated through the activation layer.
- Computing average of last *K* predictions and choosing label with maximum corresponding likelihood.
- Labeling the frame and writing output frame to the disk.
- Fastening training period with the use of the RELU (i.e. rectified linear unit).
- Throughout the training Loss layer is added at the end for giving a feedback to the NN.
- Storing the labeling and after that, retrieval video.

This is the first step in the process. A convolutional layer has a number of the filters performing the convolutional operations. Each one of the images is considered as pixel value matrix.



5.2 ReLU layer: which carries out element-wise process and assigns all the negative pixels a value of 0. It presents the non-linearity to network, and generated output is **rectified feature map**. The following is a graph of ReLU function.

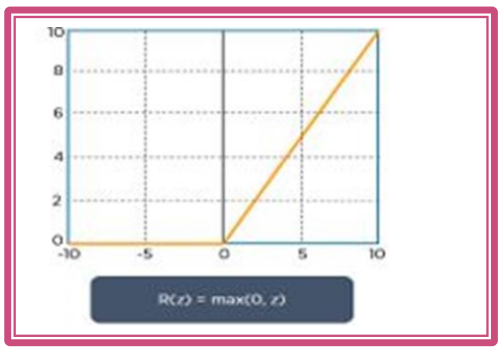


Fig.5: Scanning of the original image by multiple convolutions ReLU layer in order to locate features.

5.3 Pooling Layer

It is a down sampling process, which is utilized for reducing the feature map dimensionality. Now, ReLU map goes through some pooling layer for the generation of pooled feature map. This layer utilizes a variety of the filters for the identification of a variety of the image parts such as the corners, edges, feathers, body, beak and eyes. The following

DOI: 10.31695/IJASRE.2023.9.12.7

step in this procedure is referred to as **flattening**, which is utilized for the conversion of all resulting 2D arrays from pooled the feature maps to a one long continuous linear vector.

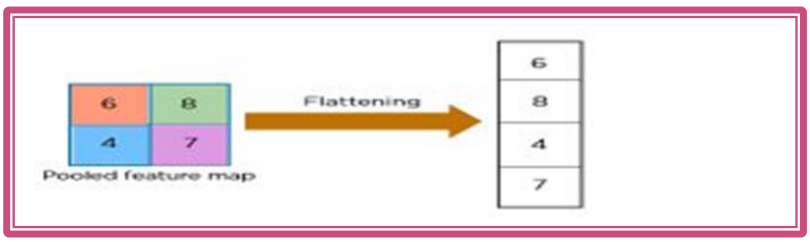


Fig .6: flattened matrix fed as an input into fully connected layer for the classification of image.

The image pixels are given into convolution layer, which then carries out convolutional process, it produces convolved map, which applied to the function of ReLU for the generation of rectified feature map. A variety of the layers of pooling with a variety of the filters have been utilized for the identification of certain image parts. After that, the pooled feature map is flattened then fed into fully connected layer for the purpose of getting final result.

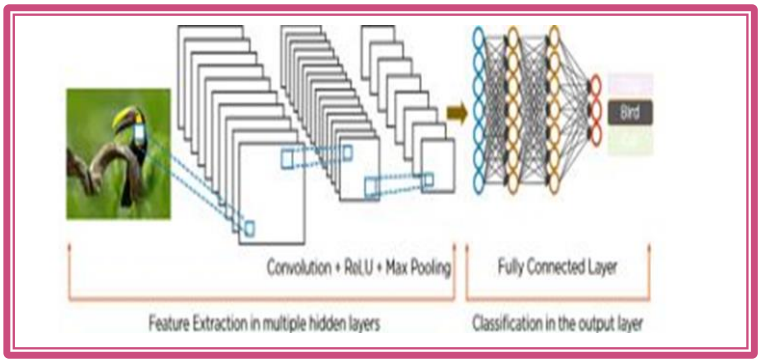


Fig. 7: CNN Layer

6. EXPERIMENTAL RESULTS

In experiment process, for the training and the testing of the CNN classifier, a data-set was utilized, consisting of 40 videos and classified to 4 classes, which have been illustrated in Figure7. There are 10 cricket videos, 10 animal videos, 10 airplane videos and 10 flower videos exhibited in Figure 8. From 40 videos, after the application of the detection of the scene changes and extraction of the key frame, 145 key frames have been obtained, which are depicted in Figure9, 32 frames have been extracted for the Animal videos, 18 for the flower videos, 34 for the cricket videos, and 61 for the Airplane videos. The performance of video retrieval is usually measured by the following two metrics:

Precision: In the field of video retrieval, **precision** is the fraction of video that are relevant to the search. A good retrieval system should only retrieve relevant items.

Precision = TP / (TP + FP) (where TP is True Positives and FP is False Positives)

Recall: in video retrieval is the fraction of the documents that are relevant to the query that are successfully retrieved. A good retrieval system should retrieve as many relevant items as possible.

Recall= TP/ (TP+FN) (where FN is False Negative)

DOI: 10.31695/IJASRE.2023.9.12.7

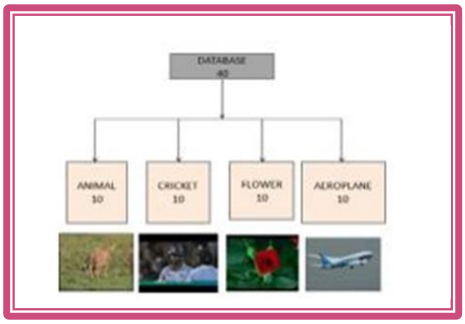


Fig.8: (Video Categories)

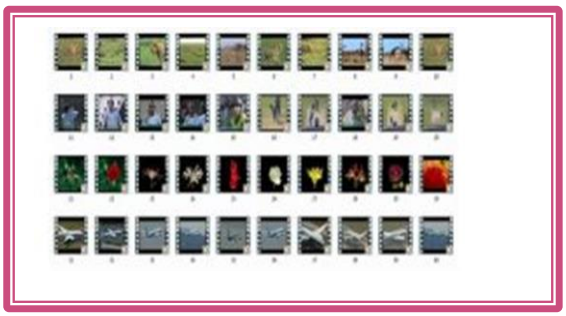


Fig .9:(Video Database)

Fig .10: 145 key frames that have been extracted for 40 data-base videos.

Fig .11: Retrieval Video for animal

Fig.12: Retrieval Video query for airplane

DOI: 10.31695/IJASRE.2023.9.12.7

Table 1 shows the experimental results of proposed video retrieval system. We calculate the result by the two metrics precision and recall.

Table 1: Experimental results

S-No	Frame	Precision	*Recall*
Animal	**32**	**1.00**	**0.083**
Cricket	**34**	**1.00**	**0.500**
Flower	**18**	**0.98**	**0.440**
Airplane	**61**	**0.96**	**0.400**

Results show that the performance of the system is more than 95%. The video retrieval is more efficient than previous approaches because it is invariant to illumination changes. Fig.13 shows the graph between Precision and Recall.

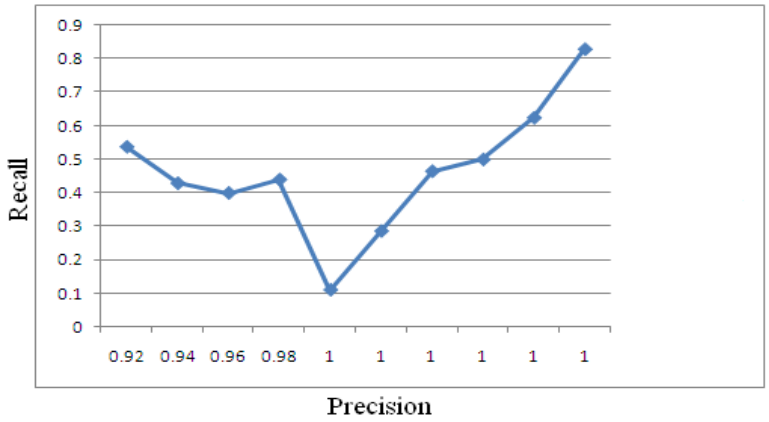


Fig.13: Graph between Precision and Recall

7. CONCLUSION

The Retrieval of the Video is a field with a wide range, integrating the features that include the AI, ML, database management system, and so on. The present study work presented an approach for the semantic concept-based retrieval of the video with the use of state-of-art classifier that has been constructed with the use of the asymmetrically trained deep CNN to deal data-set. There were many algorithms that are rooted in those areas for performing a variety of the tasks of video retrieval. In the system that has been suggested in the present study, a retrieval system has been implemented through the integration of a variety of the query video frame features. Several of the video data-sets clearly show that the presented method could achieve more sufficient performance in comparison with other approaches of video retrieval. Which has been shown by the finding that multiple features result in producing sufficient system as values of recall and precision are enhanced.

REFRENCES

1. Varsharani Babar, Shivanee Bhor, Krupa Kamble, Ms Arathi Kamble, "CONCEPT BASED VIDEO RETRIEVAL", international Journal of Emerging Technologies and Innovative Research(JETIR), ISSN:(2349-5162), Vol. (6), Issue (6), PP:(203-205), 2019. https://www.jetir.org/papers/JETIR1907029.pdf.
2. El Mehdi Saoudi, Said Jai-Andoloussi, "A distributed Content –Based Video Retrieval system foe large datasets", (2021), Journal of Big Data, 8, no. (87). https://doi.org/10.1186/s40537-021- 00479-x.
3. Matheel E. Abdulmunem , Eman Hato " Semantic Based Video Retrieval System: Survey ",Iraqi Journal of Science , EISSN:(2312-1637), ISSN: (0067- 2904), Vol. (59), No. (2A), 2018. https://ijs.uobaghdad.edu.iq/index.php/eijs/article/view/26.
4. V. A. Wankhede and P. S. Mohod, "Content-based image retrieval from videos using CBIR and ABIR algorithm," (2015) Global Conference on Communication Technologies (GCCT), pp. (767-771). **doi:** 10.1109/GCCT.2015.7342767

DOI: 10.31695/IJASRE.2023.9.12.7

5. Asif Ansari, Muzammil H Mohammed, "Content based Video Retrieval Systems-Methods Techniques Trends and Challenges", (2015), International Journal of Computer Applications, vol. (112), No. (7), PP:(13- 23). https://www.ijcaonline.org/archives/volume112/number7/19678-1402.
6. Reem A.K. Aljorani Boshra F. Zopon Al_Bayaty," On Demand Video Retrieval Based on Arabic TEXT", (2021), Turkish Journal of Computer and Mathematics Education (TURCOMAT), e-ISSN (1309-4653),Vol. (12) No. (6), PP:(3902-3912). https://www.turcomat.org/index.php/turkbilmat/article/view/7856.
7. LeCun, Y., Bengio, Y. & Hinton, G. "Deep learning" Nature 521, (436–444),(2015). https://doi.org/10.1038/nature14539.
8. Zhu, M. , He, Y. and He, Q. "A Review of Researches on Deep Learning in Remote Sensing Application", (2019), International Journal of Geosciences, ISSN Print:(2156-8359), ISSN online: (2156-8367), vol. (10), no. (1). doi: 10.4236/ijg.2019.101001.
9. Haohan Wang and Bhiksha Raj," On the Origin of Deep Learning" (2017) https://arxiv.org/pdf/1702.07800.
10. Zhuang Wu, Shanshan Jiang, Xiao lei Zhou,, et al. "Application of Image retrieval based on convolutional neural networks and Hu invariant moment algorithm in computer telecommunications "computer-communications, vol.(150) , January(2020) , PP (729-738). https://doi.org/10.1016/j.comcom.2019.11.053.
11. X. Xing, Y. Gui, C. Dai and J. S. Liu, "NGM, "Neural Gaussian Mirror for Controlled Feature Selection in Neural Networks," (2020),19th IEEE International Conference on Machine Learning and Applications (ICMLA), pp. (148-152).doi: **10.1109/ICMLA51294.2020.00032.**
12. Xavier-Glorot, Antoine-Bordes and Y-Bengio "Deep Sparse RectifierNetworks"2010, 2010),Conference Paper in Journal of Machine Learning Research. https://www.researchgate.net/publication/215616967_Deep_Sparse_Rectifier_Neural_Networks.
13. George E. Dahl, Tara N. Sainath, Geoffrey E. Hinton "Improving deep neural networks for LVCSR using rectified linear units and dropout,"(2013), IEEE International Conference on Acoustics, Speech and Signal Processing, pp. (8609-8613). doi: 10.1109/ICASSP.2013.6639346
14. Yajiao Dond, Jianguo Li "Video retrieval based on deep convolutionalneural network"., (ICMSSP '18: Proceedings of the 3rd International Conference on Multimedia Systems and Signal Processing (2018), Pages (12–16). https://dl.acm.org/doi/10.1145/3220162.3220168.
15. Yu-Fei Ma, Hong-Jiang Zhang, "Motion Pattern-Based Video Classification and Retrieval", (2003), EURASIP Journal on Advances in Single Processing, (141352) Vol. (2), pp:(199–208). https://doi.org/10.1155/S1110865703211021.
16. Wu, Yang, Zheng, Zhou, "Motion sketch based crowd video retrieval", (2017). Multimedia Tools and Applications, 76, (20167- 20195). https://doi.org/10.1007/s11042-017-4568-2.
17. M. Aharon, M. Elad and A. Bruckstein, "K-SVD" An algorithm for designing over complete dictionaries for sparse representation,"(2006), in IEEE Transactions on Signal *Processing*, vol. (54), Issue. (11), pp. (4311-4322). **doi:** 10.1109/TSP.2006.881199.
18. Lu, H Li, in Conference of the North American Chapter of the Association for Computational Linguistics: Tutorial. Recent progress in deep learning for NLP (2016), pp. (1113). **doi:**10.18653/v1/N16-4004.
19. Yoon Kim, "Convolutional neural networks for sentence classification, "Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing Association for Computational Linguistics. pp: (1746-1751). **doi:**10.3115/v1/D14-1181.
20 .Nal Kalchbrenner, Edward Grefenstette,Phil Blunsom,"A convolutional neural network for modellingsentences"(2014),arXiv:1404.2188v1[cs.CL]. https://doi.org/10.48550/arXiv.1404.2188
21. Gautam Pal, Dwijen Rudrapaul, Suvojit Acharjee, Ruben Ray, Sayan Chakraborty, Nilanjan Dey." Video shot boundary detection" (2015), Emerging ICT for Bridging the Future Vol (2) Advances in Intelligent Systems and Computing 338.https://doi.org/10.1007/978-3-319-13731-5_14
22. Miggi Zwicklbauer, Willy Lamm Martin Gordon, et al.," Video Analysis for Interactive Story Creation: TheSandmännchen Showcase", Proc. AI4TV Workshop @ ACM MM (2020). PP: (17-24). https://doi.org/10.1145/3422839.3423061

DOI: 10.31695/IJASRE.2023.9.12.7

23. D. G. Lowe, "Object recognition from local scale-invariant features, Proceedings of the Seventh IEEE International Conference on Computer Vision, *(*1999), pp. (1150-1157), vol. (2).
doi: 10.1109/ICCV.1999.790410

2**4.** Sneha Pai, Ramesha Shettigar, "Accuracy Analysis of SIFT and SURF Descriptor based on Gender Classification" INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) ISSN: (2278-0181), RTESIT (2019) Conference Proceedings, vol. (7), Issue (8). https://www.ijert.org/Accuracy-Analysis- of-SIFT-and-SURF-Descriptor-based-on-Gender-Classification.

25. D. Willy, A. Noviyanto and A. M. Arymurthy, "Evaluation of SIFT and SURF features in the songket recognition*",* (2013*),* International Conference on Advanced Computer Science and Information Systems (ICACSIS), pp. (393-396). doi**:** 10.1109/ICACSIS.2013.6761607.

26. A. Ullah, K. Muhammad, T. Hussain, S. W. Baik and V. H. C. De Albuquerque, "Event-Oriented 3D Convolutional Features Selection and Hash Codes Generation Using PCA for Video Retrieval," in IEEE Access, vol. (8), pp. (196529-196540), (2020).doi: **10.1109/ACCESS.2020.3029834.**

27. Ryfial Azhar, Desmin Tuwohingide, Dasrit Kamudi, Sarimuddin, Nanik Suciati,," Batik Image Classification Using SIFT Feature Extraction" Procedia Computer Science 72 (2015) 24 – 30 https://doi.org/10.1016/j.procs.2015.12.101

28. Zong-Yan Li, Li-mei Song, Jiang-tao Xi, et al. "A stereo matching algorithm based on SIFT feature and homography matrix,". optoelectronics Letters, vol. (11), no. (5), pp. (390–394),2015. https://doi.org/10.1007/s11801-015-5146-3.

29. Dong Haifeng and Yao Jun, "Research on robot binocular rangingbased on SIFT feature extraction algorithm," Journal of Physics: Conference Series, vol. (1607), no. (1), pp. (012015–012021), (2020).https://doi.org/10.1088/1742-6596/1607/1/012015.

30. Lingqiang Kong**, "**SIFT Feature-Based Video Camera Boundary Detection Algorithm" (2021), Complexity, Volume (2021), Article ID (5587873), pp. (11). https://doi.org/10.1155/2021/5587873.

31. Chhabra, P., Garg, N.K. & Kumar, M. Content-based image retrieval system using ORB and SIFT features. *Neu ral Comput & Applic* **32**, 2725–2733 (2020). https://doi.org/10.1007/s00521-018-3677-9.

32. T. A. Al-Shurbaji, K. A. AlKaabneh, I. Alhadid and R. Masa'deh, "An optimized scale-invariant feature transform using chamfer distance in image matching," Intelligent Automation & Soft Computing, vol. 31, no.2, pp. 971–985, 2022.https://doi.org/10.32604/iasc.2022.019654.

DOI: 10.31695/IJASRE.2023.9.12.7