Graph Based Video Sequence Matching & BoF Method for Video Copy detection

Kakade Avinash Hanumant
PG Student,
Department of Computer Engineering.
G.H. Raisoni C. O. E. & Management,
Ahmednagar., India
E-mail: avinash.kakade12@gmail.com

Prof. Mansi Bhonsle
Assistant Professor,
G.H. Raisoni C. O. E. & Management,
Wagholi, Dist:-Pune, India,
Email-id:mansi.bhonsle@gmail.com

Abstract— In this paper we propose video copy detection method using Bag-of-Features and showing acyclic graph of matching frames of videos. This include use of both local (line, texture, color) and global (Scale Invariant Feature Transform i.e. SIFT) features. This process includes dividing video into small frames using dual threshold method which eliminates the redundant frames and select unique key frames. After that from each key frame binary features are extracted which known as Bag of Features (BoF) which are get stored into the database in format of matrix. When any query video is being uploading, same features are extracted and compared with stored database to detect copied video. If video detected as copied then using Graph Based Sequence Matching Method, actual matched sequence between key frames is displayed in acyclic graph.

Keywords—*SIFT Features, CBCD, Bag of Features (BoF), Dual Threshold Method, Graph Based Sequence Matching.*

*****

I. INTRODUCTION

Video copy detection techniques works according to their content i.e. images. These techniques are known as CBCD (Content Based Copy Detection) techniques. Comparing two different videos with each other includes comparing features of videos. There are two types of features are extracted from video key frames [7] first is, Local features which are basically description of line, texture, color of image. Second is global feature which are combination of scale rotation of images these features are known as Scale Invariant Feature Transform i.e. SIFT Features. [1] According to these properties of features, copy detection of copied video is become easier. But there are also challenging task in video copy detection like detection of change in rotation, Scale, video in video etc. To achieve this challenge there are different types of CBCD techniques like Singular Value Decomposition (SVD) [5]. All those methods are working according to contents of video for e.g. images, features value etc. Video copy detection techniques require the stored database of binary features. After feature extraction from frames they are stored into database and used to match the features of query video. Storage of binary features includes structures like Indexing, Clustering [8], Matrix format etc. Video copy detection is also possible to detect the face and regions, motions, visual cues [2], Spatial-Temporal images [3], within video. In such techniques Temporal Informative Images are extracted from the video which are known as TIRI images [8]. These images basically show the actual motion of particular image or frame. Due to selection of such spatial images, redundancy is avoided and large number of same images gets reduced.

Implementation this technique is divided into two parts: *1. An Offline Process.* This includes the extraction of features from the key frames of the video from reference video database. Key frames are selected from the video by dividing it into small frames using Dual Threshold Method [1] which also rejects the redundant frames. After that, binary features are extracted from key frames and stored into the data structure. These features are robust and unique. This database is used to compare the features of query video while uploading on any type of server.

2. An online process. This is part of detection of copied video at real time on web server. Key frames are get extracted from the query video same as described in offline process and also binary features are gathered from these key frames. After getting these binary features they are compared with features stored into database to detect the copy. Because of robustness and uniqueness of these features exact match between video is detected. In proposed method it also detects the change in transformation, scale and rotations. Due to this, providing copy right protection to videos is not needed and robust security is provided towards the videos over the web servers which contains the authorized videos which should be kept protected. In this paper, we focus on detecting video copy by comparing Bag of Features of the videos and generating video sequence matching acyclic graph which shows the matched sequence of the key frames of the video. It is able to detect the transformation, scale, color change in video and also detects the picture in picture. Graph based sequence matching method shows exact matched picture within picture in format of acyclic graph.

II. RELATED WORK

Lot of methods is proposed for video copy detection. In 2010 new method is implemented to detect the facial and activity based copy detection [2]. That technique used the detection process into three different types. First types was based on the facial shot matching detection in which according to extended

 ISSN: 2321-8169
3844 - 3848

body regions it detects the same images of particular person. Second type focused on detection of the video on the basis of activity which detects the exact activity matched between two video clips. Third type was able to detects the copied video with non facial matching results also which was basically uses the low level MPEG-7 descriptor and similarity calculation with respect to dynamic-weighted features [2]. But these techniques had disadvantages that it was unable to detects the changes in video with respect to the scale, variation, color changed Robust video copy detection needs the fast result retrieval from the managed database. To point of view these issues cluster based similarity search [8] and inverted file structure was innovated. These techniques focused on video copy detection using extraction of the Temporal Informative Images (TIRI) [8]. These images denote the particular motion of the video and combination of the multiple images. Fast data retrieval was achieved by cluster based similarity search which basically works on the binary fingerprints extracted from the images of the query video. These binary fingerprints are robust and unique and also same in size for example 16bit.

This technique divides these fingerprints into the block and calculates the cluster head from that fingerprint. That fingerprint is attached to the head into database. Because of this head calculation fingerprint matching result was fast.

Some techniques were developed for detection of spatial temporal consistency [3] of video frames after matching individual frames. This technique searches the strongly deformed videos in relatively small datasets. Instead of extracting features, this technique extract the local signature from the video frames and inverted file structure is applied to fast retrieval of the local signatures from the database. One of the advantages of this technique is that retrieval of data from database is fast due to inverted file structure. But limitation of this technique is, this works over only individual frames which are same with spatial movement. This technique is not able to work with large size of videos. Using local spatial temporal features [4] detecting copied video is also works fast. This technique clearly uses the local spatial temporal features instead of purely spatial images and due to this its results into robustness. This technique also starts with dividing the video into frames and selects the temporal similar key frames. For database storage this technique uses the Locality Secure Hash (LSH) [4] technique. One of the big advantages of this technique is that, it detects the high dimensional similarities from the videos. But similarities between two videos are not detected by this method. This is limitation and new challenge to this technique

This method requires only two spatially same key frames to compare each other and due to this, method detects only two exactly same videos not similarities between different videos. Previously described all methods were works only for local features which are easier to copy detection. But it's also challenge to work for both local and global features. Singular Value Decomposition (SVD) [5] is such a technique which works using SIFT (Scale Invariant Feature Transform) [5]. SIFT features are combination of local and global features. Using this SIFT features it's possible to detects the copied videos with respect to change in scale, variation in scale and rotation etc. It's also possible to detect. the video in video or picture in picture.

III. PROPOSED SYSTEM

Main flow of proposed method follows into three parts.

1. Extracting features: i) Dual Threshold Method.
   ii) Bag of Feature (BoF) extraction.
   iii) Feature Database.
2. Feature Comparison.
3. Graph Based Sequence Matching Method.

Proposed technique works in two phases one is offline and second is online phase. Basically there is need to maintain the sample feature database from the videos, and this process is accomplished in offline phase.

Online phase is actual matching of the query video and target video in which feature extraction is also applied and get compared with stored database. At last graph of matched sequences between frames are shown. Both offline and online phases are shown in the following fig 1.

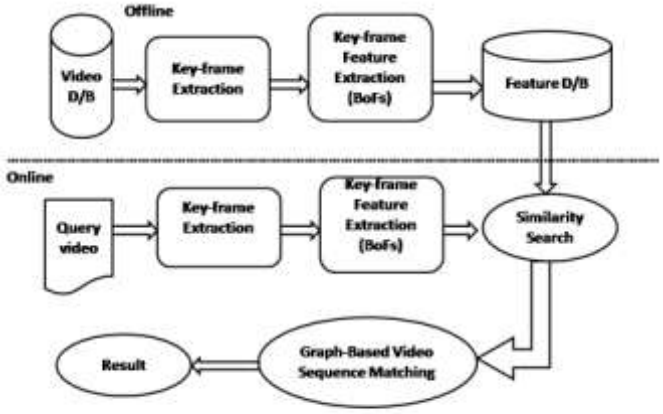


Fig: 1) Flow of System.

1. Extracting Features: Main part of the proposed method is extraction of binary features from the video. These features are stored into database to compare with query video. First step to extract feature is divide video into frames. To divide video into frames dual threshold method is used.

i). Dual Threshold Method:

In first phase there is need to divide video into the frames. After dividing video into frames we get large numbers of frames. But for time consistency there is need to eliminate the redundant frames. Dual threshold method eliminates the redundant frames from all extracted frames; this process is nothing but selection of the key frame. Key frames describe the actual motion or action performs in that set of key frames (Segments). All frames are divided into segments i.e. combination of frames. From each segment one key frame is selected as shown in fig.2.

 ISSN: 2321-8169
3844 - 3848

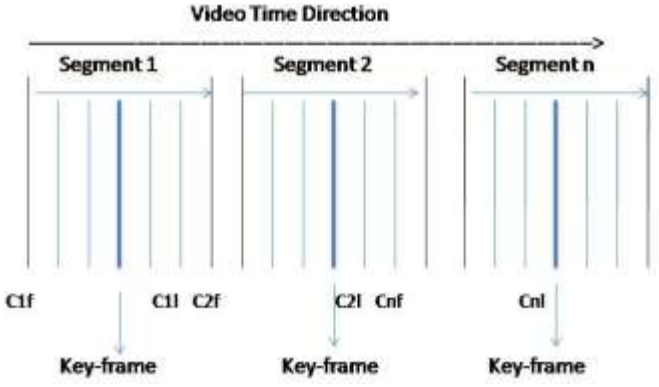


Fig. 2 Dual Threshold Method

ii). Bag of Features Extraction:

Features are nothing but local and global descriptor values. Features are binary values extracted from the images/frames. We are maintaining local features like Scale Invariant Feature Transform (SIFT) which is description of line, color, texture values of the key frames. To detect copied video robustly there is need to extract both local and global features. In this technique we are extracting Bag of Visual Words from each key frame and storing into the database in fix dimensional format.

iii). Features Database:

Extracted features are retrieved in the matrix format 128×128 dimensional. For fast matching result we need to store these features into database with simple format. For that purpose here we are using the K means clustering algorithm to get exact K size clusters of feature to store into database. Due to clusters data retrieval is become easier. These fix size of features are stored into the database with fix dimensional size. Because of fix size clusters comparison operation is simple to work.

2. Features Comparison.

When any user uploading the video, the feature extraction is applied over it, and after getting features these are get compared with already stored features in database to detect the video is copied or not.

3. Graph Based Sequence Matching Method.

If query video and target video is matched, then sequence matched between key frames is possible to show using the Graph Based Sequence Matching Method.
Graph Based Sequence Matching Method works into two phases in which first phase is same till the key frame selection from the videos
In second phase matching of frames of query video and target video is performed. For that purpose KNN similarity algorithm is used, which retrieves the similar frames of query and target video. KNN algorithm applies the $sim\ (C_i^Q, C_j^T)$ function over segments, which returns the K number of matched segments. Where C_i^Q is i^{th} frame of query video and C_j^T is the jth frame of target video.

This similarity matched result also converted in graph. Consider M_{ij} represents the match between C_i^Q and C_j^T. Then to satisfy edge between this vertexes following condition should matched. $(i\text{-}l) * (j\text{-}m)>0$, then M_{ij} and M_{lm} satisfies time direction consistency.
Sequence matched graph is shown in following Fig.3 In which M2 and M3 not satisfying time consistency because its not satisfying condition. i.e. (2-3) *(29-26)<0 which should be greater than zero.

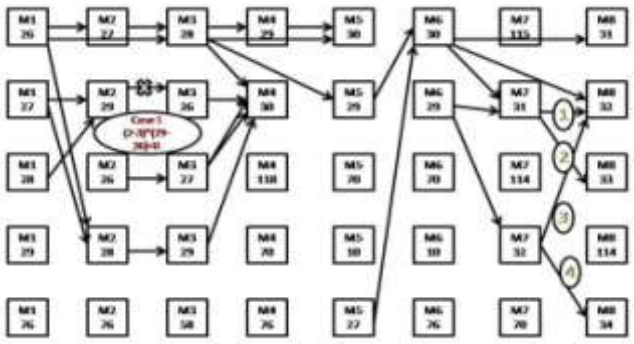


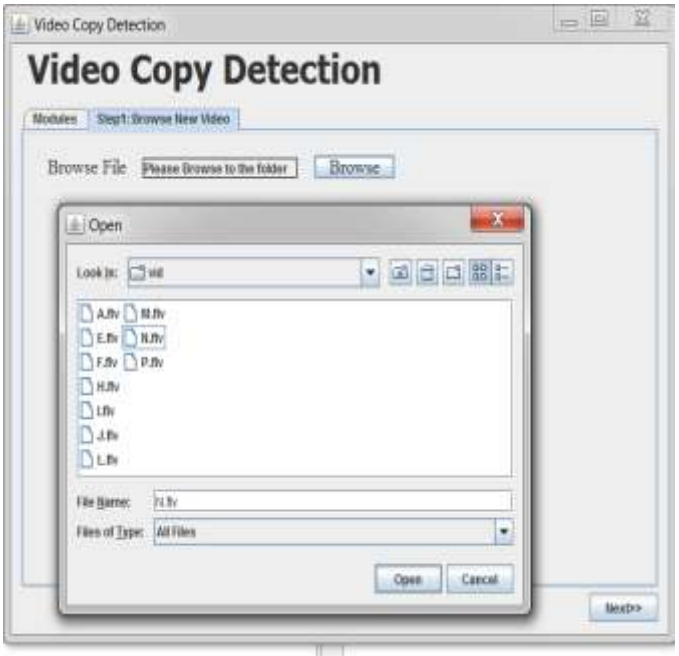
Fig.3) Graph Based Sequence Matching Method

IV. DATASET

Dataset
Small size videos are available at following website Final CBCD Evaluation Plan TRECVID 2010 (V2), http://www-nlpir.nist.gov/projects/tv2010/Evaluation-cbcd-v1.3. tml#eval, 2010.Also user can add their different videos to detect copies.
Dual threshold method is applied over these videos to devides video into key frames. Features extracted from the key frames are stored into database into fix size using clustering algorithm.

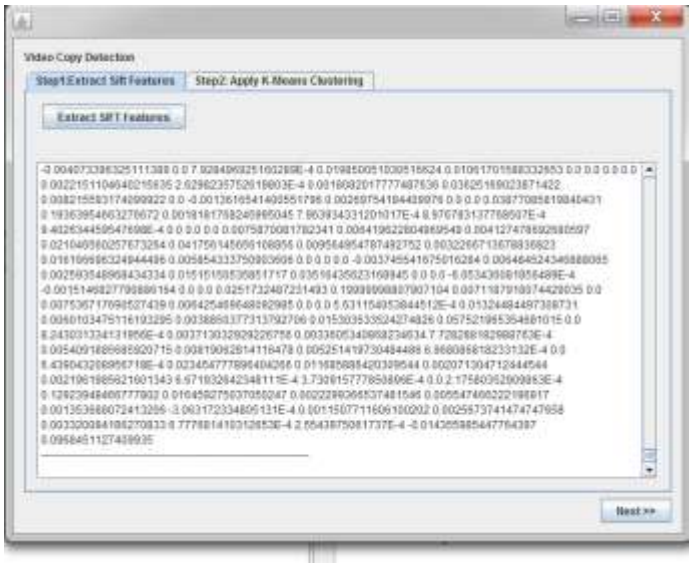
V. RESULTS

1: Selecting Video to Upload

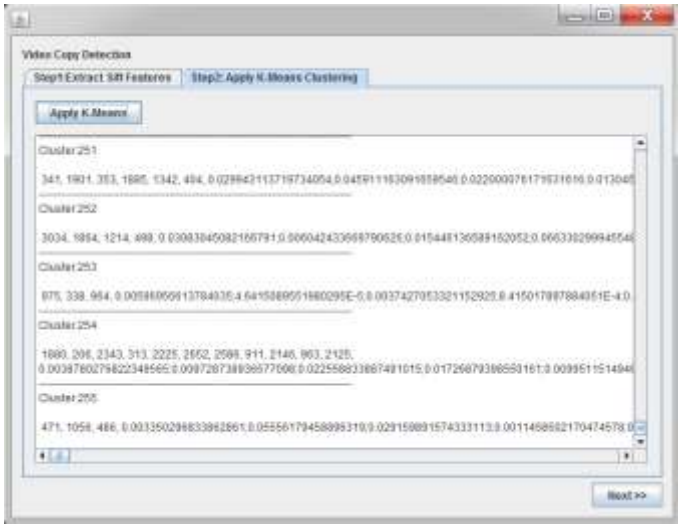


 ISSN: 2321-8169
3844 - 3848

2: Extracting SIFT Features from frames of Video.

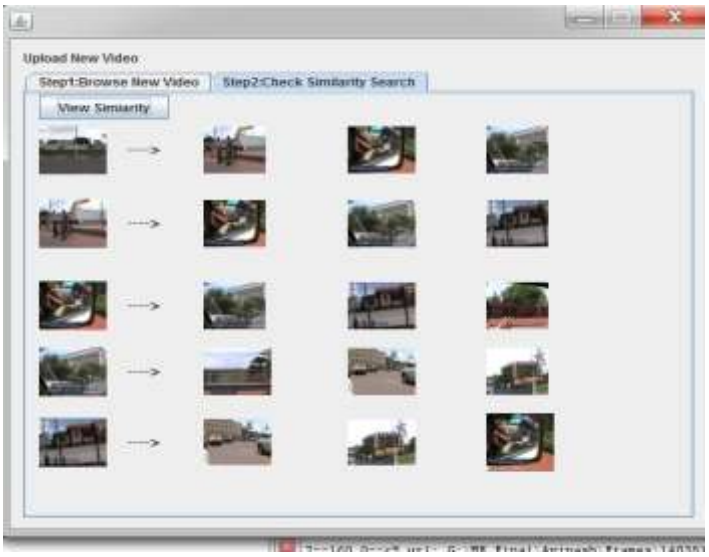


3: Applying Clustering on Features.



4: After Comparison Duplicate video is detected.

5: Showing Similarity Graph of Images .



VI.CONCLUSION

The proposed method presented new CBCD video copy detection technique with Bag of Features i.e. BOF.
Basically proposed method woks over Bag of Features i.e. SIFT features of key frames extracted from the video. Method results into detection with respect to change in scale, rotation, color etc. Feature retrieval is also fast due to the use of matrix database storage binary Bag of visual words. Graph based sequence matching shows exact match between frames of stored and query video.

VII. ACKNOWLEDGMENT

I express great many thanks to Prof. Mansi Bhonsle for her great effort of supervising and leading me, to accomplish this fine work. Also to college and department staff, they were a great source of support and encouragement. To my friends and family, for their warm, kind encourages and loves. To every person gave us something too light my pathway, I thanks for believing in me.

REFERENCES

[1] Hong Liu, Hong Lu, Member, IEEE, and Xiangyang Xue, Member, IEEE, “A Segmentation and Graph-Based Video Sequence Matching Method for Video Copy Detection” IEEE Transactions on Knowledge and Data Engineering, vol. 25, no. 8, August 2013

[2] O. Ku¨cu¨ ktunc, M. Bastan, U. Gu¨du¨ kbay, and O ¨ .Ulusoy, “Video Copy Detection Using Multiple Visual Cues and MPEG-7

Descriptors,” J. Visual Comm. Image Representation, vol. 21,pp. 838-849, 2010.

[3] M. Douze, H. Je´gou, and C. Schmid, “An Image-Based Approach to Video Copy Detection with Spatio-Temporal Post-Filtering,” IEEE Trans. Multimedia, vol. 12, no. 4, pp. 257-266, June 2010.

[4] G. Willems, T. Tuytelaars, and L.V. Gool, “Spatio-TemporalFeatures for Robust Content-Based Video Copy Detection,” Proc.ACM Int’l Conf. Multimedia Information Retrieval (MIR), pp. 283- 290, 2008.

ISSN: 2321-8169
3844 - 3848

[5] E. Delponte, F. Isgro` , F. Odone, and A. Verri, “SVD-Matching Using Sift Features,” Graphical Models, vol. 68, no. 5, pp. 415-431, 2006

[6] F. Dufaux, “Key Frame Selection to Represent a Video,”Proc. IEEE Int’l Conf. Image Processing, vol. 2, pp. 275-278, 2000”.

[7] TREC Video Retrieval Evaluation, http://www nlpir.nist.gov/ projects/t01v/, 2006.

[8] Final CBCD Evaluation Plan TRECVID 2010 (V2), http://

[9] www-nlpir.nist.gov/projects/tv2010/Evaluation-cbcd-v1.3.htm#eval, 2010

[10] H. Jegou, M. Douze, and C. Schmid, “Hamming Embedding and Weak Geometric Consistency for Large Scale Image Search,” Proc. European Conf. Computer Vision, 2008.