Detecting Video Sequence Matching Using Segmentation Method

[1]K.Girija, [2]S.Herman Jeeva, [3]M.Soniya, [4]P.Sabarinathan

[1,2,3] PG scholar, Department of Computer Science and Engineering, Pavendar Bharathidasan college of Engineering and Technology, Tiruchirappalli, Tamilnadu, India

[4] Assistant Professor, Department of Computer Science and Engineering, Pavendar Bharathidasan college of Engineering and Technology, Tiruchirappalli, Tamilnadu, India.

ABSTRACT - There are a number of methods available for video copy detection. Some of the methods were employing the application of local and global descriptors which were found to be ineffective in detections involving complex transformations. In order to overcome the above specified inefficiency, Scale Invariant Feature Transform (SIFT) descriptor came into picture but was found to have a high computational cost. The method proposed in this paper involving five different types of MPEG-7 descriptors namely Color and Edge Directivity Descriptor (CEDD), Fuzzy Color and Texture Histogram (FCTH), Scalable Color Descriptor (SCD), Edge Histogram Descriptor (EHD), Color Layout Descriptor (CLD) for extracting the features of the frames in the selected video is found to be cost effective and efficient even in case of high level of transformations. This paper also throws light on certain improvements in graph-based video sequence matching method which is used to overcome the level of noise, to detect videos with different frame rates and optimal sequence matching is found automatically from the disordered video sequences by applying spatial features during copy detection. Experimental results have showed that the proposed method is far effective than the previously existing video detection scenarios.

KEYWORDS -Video copy detection; graph; SIFT feature; SVD; CEDD, FCTH, SCD, EHD, CLD descriptors; graph-based matching.

I. INTRODUCTION

Now-a-days, server space is becoming a major issue for high level organizations to maintain enormous amount of data. For instance, organizations like YouTube, Google, Metacafe and others dealing with enormous video storage are found to have acres of racks consisting of hard disks each could be holding a capacity of around 1 to 2TB. But we are aware of the fact that most of the videos available in these video storage websites are redundant. According to the recent statistics, there are about 27 percent of redundant videos in You Tube, Google videos. Redundant videos are of two types: Copy videos and near duplicate videos.

A copy can be defined as a segment of video derived from another video, usually undergone a lot of transformations, such as Cam-cording, PiP (Picture in Picture), Insertions of patterns such as captions, subtitles, logo; Strong re-encoding; Change of gamma; Decrease in quality: Blur, frame dropping, compression, ratio and white noise, Post production is shown in TABLE 1. All of the above specified transformations can be done to a specific original video or else some of these transformations can also be done based on the needs of the one who is implementing these transformations. Both of these above specified cases are considered to be a copy of the original version of that respective video which is available in the dataset maintained by us.

Near duplicate videos are the ones representing the same sequence of actions or an event which are recorded by two different cameras from different position or angle. Even though these two videos are representing the same sequence of actions, they are not considered to be copies since they are not edited from an existing original video recorded by someone else. So this paper is not concerned about detecting near duplicate videos.

The ultimate goal of video copy detection is to decide whether a video query is copied from a video available in the video dataset. A copy could have undergone various transformations specified earlier. If the system finds the sequence matching results to be the same in the client and server side, the system would prompt with a message saying that the video input of the user is a copied one whose original video is already available in the server.

The framework of video copy detection is done as two parts:
(1) Client side: Key frames are extracted from the reference video dataset and features are extracted from those key frames. The features extracted should not be vulnerable to video transformation. The features can be stored in a feature database to make similarity comparison effective.
(2) Server side: Query videos are verified. Features are extracted from the videos and are compared with the features which are already stored in the database. The sequence matching results are returned along with the detection results [5].

Based on the study, the transformations involving picture in picture is complicated to be detected [9], [10]. And for detecting such video copies, local feature of SIFT is found to be normally valid. However, matching which is done based on the local features in certain cases involving two videos seems to have a high level of computational complexity.

TABLE-I

VARIOUS TYPES OF TRANSFORMATIONS
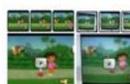
Methods to detect picture in picture and proposing graph-based sequence matching method are focused here.

II. RELATED WORK

Video copy detection was a problem being handled over a long period of time particularly when the videos are heavily transformed. A lot of research has been done in this area. There was no shot boundary detection algorithm available specialized for Content Based Copy Detection termed as CBCD. So Onur Kucuktunc, Ugur Gudukbay and Ozgur U have proposed an automatic shot boundary detection algorithm meant for videos with heavy transformations applied which is based on Fuzzy color histogram. This method is efficient in the way that we would consider not only the values of perfect true and perfect false but also the values of partial true and false. Apart from this, other problems were copyright violations and image forgery. Although digital watermarking was available to save the day, it was very

difficult to design and so Yan Ke, Rahul Sukthankar and Larry H have proposed a copy detection technique using scale and rotation invariant interest point detectors.

The efficiency of this method lies in the efficient layout of data on disk. Other issues include representing video shots with the help of histograms which does not represent the spatial information about the pixels and so K. Sze, K. Lam, and G. Qiu have proposed an optimal representation of video shots representing both spatial and global information based on probability. This method has a greater computational complexity compared to alpha trimmed average histogram method. Video similarity detection using visual similarities of video clips is another issue to be dealt with since watermarking is fragile to visual transformations such as cam-coding. Onur Kucuktunc, Muhammet Bastan, Ugur Gudukbay and Ozgur U have proposed an video copy detection framework consisting of three methods namely facial shot matching along with activity subsequence matching and non-facial shot matching [10]. This method is in need of greater potential to detect copies among videos undergone cam-coding or PiP (Picture in Picture) kind of transformations.

A video sequence identification method to find a similar content to a short query clip from a long video sequence was required and so Heng Tao Shen, Jie Shao, Zi Huang and Xiao fang Z have proposed an approach based on graph transformation and matching which can even identify different ordering of frames and content editing. A novel batch query algorithm was also been proposed to retrieve similar frames. The merit of this method is that there is no need of pre-segmentation of videos as we frame sub-sampling to identify ambiguous shot boundaries [4].

III. THE PROBLEM

Watermarking is one of the classical techniques handled down for a long time for the purpose of representing the ownership of a particular video by means of inserting the logo or any other digital signature of the owner [1]. This information can be either visible to human eye or even can be hidden. This seems to be a better way for representing the ownership but it is vulnerable to the slightest changes. For instance, this watermarking information can be removed easily by anyone by making any small transformation to the watermarked video which includes re-encoding and change of bit rate. It is important to note that watermarking is not destined to retrieve a query video clip.

To overcome this demerit, CBCD (Content Based Copy Detection) came into existence. Methods which are employing the use of global descriptors were found to take the low level features into account always for the purpose of video copy detection. But this technique is found to work well only for those videos which have undergone some small amount of transformations. In order to overcome the demerits of global descriptor, local descriptor were employed for video copy detection which makes use of the local spatiotemporal features and then making use of the features around the local features which helps them to be effective even in the case of high level transformations [3]. One of the efficient local descriptor that came into existence was SIFT.

In existing system, SIFT descriptor [6] seems to have a good stability to describe video characteristics. The SIFT descriptor (Local Descriptor) is found to give a better performance in identifying the objects is shown in Figure 1. The methods based on local descriptors on points, lines and also shape found to play an important role in video copy detection.
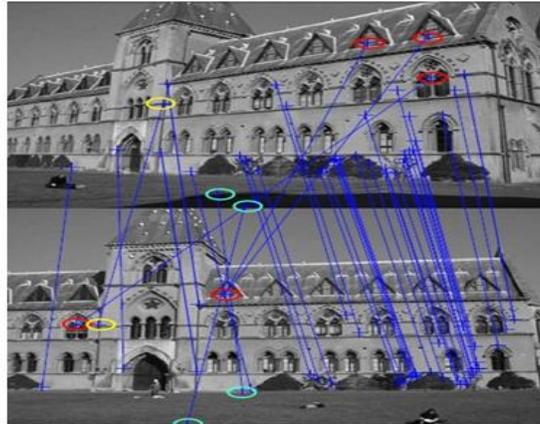
Figure 1: The Matching Result Using SIFT

Of the above mentioned properties, spatiotemporal interest points are found to play a vital role in classifying human actions and to detect periodic motions it has a good tolerance to scale changes, illumination variations and image rotations and also it has a greater strength against affine distortion and additive noise. Performance of methods based on local descriptor is found to be far better than the methods based on global descriptor in terms of logo insertion, shifting or cropping. In spite of the above specified merits, high computational cost is found to be highest demerit.

IV. CONTRIBUTION

In order to increase the efficiency of video copy detection and make it more effective, we have employed the use of five MPEG7 type descriptors such as CEDD, FCTH, SCD, EHD, CLD descriptors are used to extract the feature for Sequence Matching and also used to reduce the computational complexity. They describe elementary characters such as shape, color, texture or motion among others.

CEDD – Color & Edge Directivity Descriptor

Low computational power for extraction compared to others. It incorporates color and texture information in histogram. CEDD low level features that are extracted from the images can be used for indexing and also for retrieval.

FCTH – Fuzzy Color & Texture Histogram

FCTH is used for extraction of features that combines color and texture information in one histogram. This feature is appropriate for accurately retrieving images even in case of distortion such as deformation.

SCD –Scalable Color Descriptor

SCD derived from a color histogram defined in a Hue Saturation value color space with fixed color space quantization.

EHD – Edge Histogram Descriptor

EHD use global, semi-global, local histogram generated from local histogram bins to increase performance. MPEG-7 is used for the case of video or still images containing texture, color, shape and motion.

CLD – Color Layout Descriptor

CLD is designed to capture spatial distribution of color in image. It is used to describe color relation between sequence and group of images.

Graph-Based Method

Graph method can find the longest path in the frame matching-result graph considering time constraint

- It can automatically remove the noise caused by visual feature matching.
- It is adaptive to video frame rate change.

It can find the optimal sequence matching result automatically. The above specified five types of descriptors are used to extract exact individual features from the frames which are more accurate for comparison of the results. In addition to that graph based method is used to make this method an optimal one for video copy detection.

V. SYSTEM DESIGN

The system design involves the different steps involved in the proposed video copy detection method. The entire process is summarized in the Figure 2 which gives a clear cut idea about the proposed method.

Client side:

It is important to note that the client side and offline process are one and the same which can be represented in any of the descriptions. The procedure followed at the client side involves getting the user input in the form of video. First step would be extracting the key frames from the user given input video. The different features are extracted from the extracted key frames using the different type of descriptors and are stored into the reference database.

Server side:

As specified above, server side is similar to that of online process which can be said in either ways. The process done at the server side is similar to that of the process done at the client side which is the following. Frames are extracted from all the videos available in the server side and the features are already extracted using the proposed five types of descriptors and are stored individually in the server side database for future reference. It is to be noted that each time when the user is uploading a new video into the server, key frames are extracted along with the features and are updated into the database maintained by the server.

Process:

When both of the process is done, the next step would be to pick a random key frame from the reference database in the client side and compare it with all the extracted key frames in the client side reference database. Once the comparison is done, the descriptors would individually return the respective similarity matching results with the mathematical values representing the level of similarity such as the exact matches would be indicated with a 0.

The half matched or unmatched frames are represented with other non-zero values based on their level of matching. The same random frame from the client side is then compared with each and every key frame of all the videos available in the server side dataset and the similarity matching results are returned for each and every comparison. When comparing the similarity results between the client and server side, we can come to a conclusion whether the user input video is already available within the server space or not.

This is based on the scenario that similar frames would be having value to be zero and the unmatched frames would be having the non-zero values as specified earlier. Based on the results, system would prompt a message to the user saying "video already exists" in the case of a copied video or "video inserted successfully" in case of a new video.

The server also provides additional information such as the timestamp when the video was uploaded, date and whether the upload was a success or a failure and the number of attempts made by a respective user with a copied video since the server supports only the users who were registered with the server.
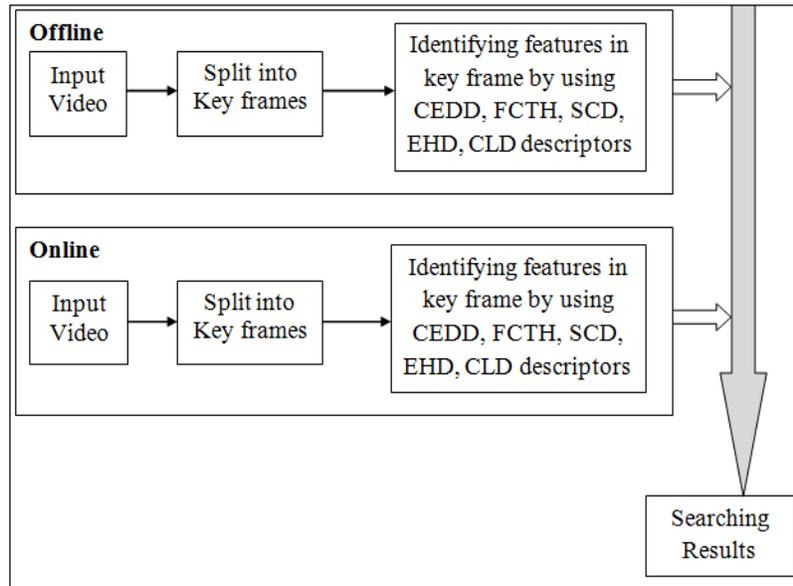
Figure 2: System Architecture Showing Client and Server Side

VI. SYSTEM IMPLEMENTATION

The above specified concepts are implemented in various steps which are described as follows:

a. Splitting of video

In order to analyze the video given as input by the user, the different desired properties of the video is found out such as color, size and quality. The input video should have high quality and brightness to achieve better results. Key frame extraction is done at this stage.

b. Performance of ROI

After key frame extraction is done, the main object features are extracted from the key frames by using ROI (Region Of Interest) and these extracted information is made used to detect the copied videos is shown in Figure 3.
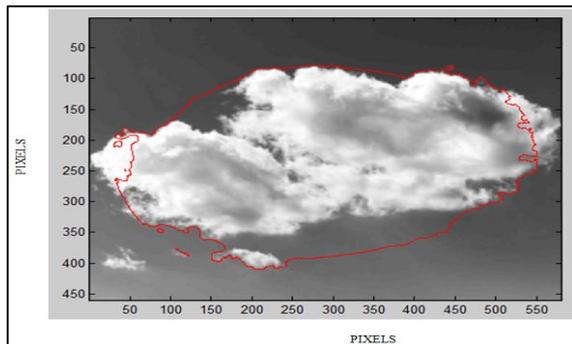

Figure 3: ROI for Extracting Main Object
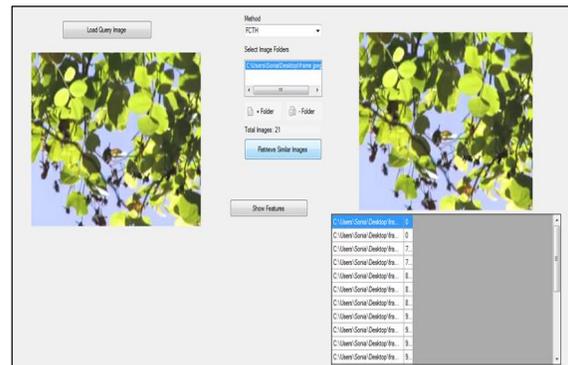
Figure 4: Identifying Sequence Matching Using FCTH Descriptors

c. RGB color histogram:

This module can used to find out various RGB properties of the frames such as mean, standard deviation, level, count at different scenarios by applying changes to the image using different types of filters such as HSL color contrast to increase the image brightness, Edge detectors, morphology, binarization and lot more.

d. Sequence Matching:

The extracted key frames are stored in the database and the random frame is chosen here to be compared with all the other frames in the dataset which is done effectively with the help of CEDD, FCTH, SCD, EHD, CLD descriptors is shown in Figure 4.

e. Similarity matching detection:

The random frame from the user input is matched with all the frames available in the server for each and every video available in the server and based on the similarity matching results, server can allow the user to upload the video if it is an original one (i.e. it is not available in the server) whereas it would not allow the user to proceed with the upload if the video is a copied one (i.e. the video is already available in the server).

VII. RESULTS AND DISCUSSIONS

The results of the compared videos are formulated in the form of graph with the help of the proposed graph based method. Results show that it is very easy to compare the results in the form of graph which shows the values formulated as a result of the comparison done between the randomly chosen frames along with each and every individual frame in the dataset. The graph also gives a clear comparative study between the values computed by all the five different type of descriptors.

TABLE-II: Similarity Matching Result of Video A

Similarity Matching Results/Descriptors	CEDD	FCTH	SCD	EHD	CLD
TEST 1	0	0	53	2.08	0
TEST 2	2.6	1.23	59	80.93	2
TEST 3	3.98	2.05	62	90.05	3.14
TEST 4	4.35	2.57	69	99.76	11.3

TABLE-III: Similarity Matching Result of Video B

Similarity Matching Results/Descriptors	CEDD	FCTH	SCD	EHD	CLD
TEST 1	0	0	63	2.15	0
TEST 2	2	1.43	72	81.73	2
TEST 3	2.73	3.22	79	84.04	2.73
TEST 4	3.44	5.41	81	115.7	15
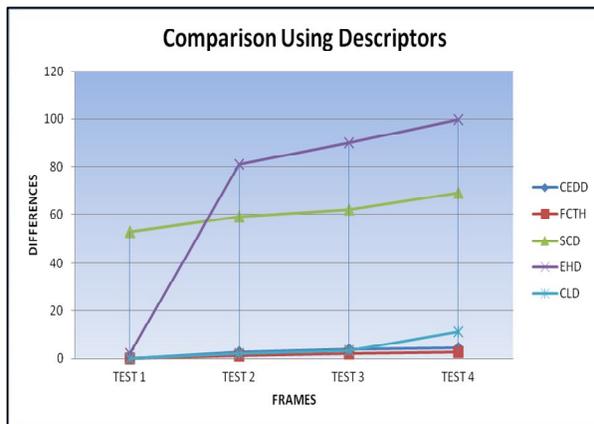
Figure 5: Graph-Based Method Representing TABLE 2
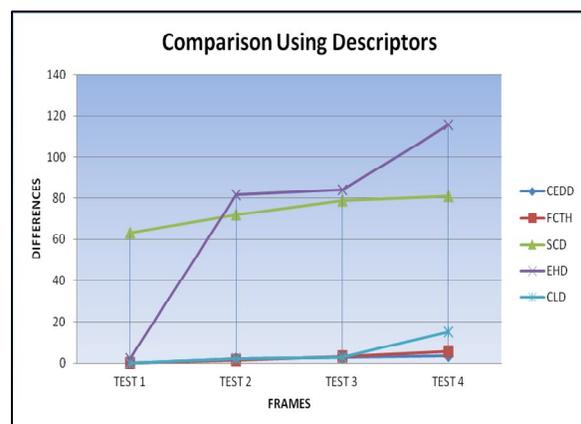
Figure 6: Graph-Based Method Representing TABLE 3

TABLE 2 shows the similarity matching result of video A done by five different type of descriptors which is represented in the form of graph in Figure 5. whereas TABLE 3 shows the similarity matching result computed for video B by the descriptors which are graphically represented in Figure 6.

On comparing the two graphs, we can come to a conclusion that the values computed by the descriptors are entirely different so that both the videos are not the same or we can say that video A is not a copy of video B. Analyzing the above results, the server would allow the user to upload video A into the server as it is not available within the server space.

VIII. CONCLUSION AND FUTURE ENHANCEMENT

Video copies are analyzed and the features are used for copy detection. Based on the analysis, local feature of SIFT was used for describing video frames. Since large number of SIFT points are extracted from a video, the copy detection involving SIFT features has high computational cost. And so the new descriptors have been introduced: EHD, CLD, SCD, CEDD, and FCTH.

Edge Histogram Descriptor (EHD) captures the edge spatial distribution around the image in five orientations: horizontal, vertical, 45° degrees, 135° degrees and non-directional. Color Layout Descriptor (CLD) is used for capturing the spatial distribution of color in an image. Scalable Color Descriptor (SCD) can obtain good performance in case of retrieval accuracy and fast matching of the image. Color and Edge Directivity Descriptor (CEDD) captures and relates shape, texture and color from an image. Fuzzy Color and Texture Histogram (FCTH) aims at capturing the image texture, shape and color.

By using these descriptors, the computational cost will be decreased and it can be used for obtaining the effectiveness in key frame feature extraction. Experimental results prove that the proposed method is found to have an optimal tradeoff between the effectiveness and efficiency of video copy detection.

Futuristic approach can be done to improve the efficiency of the descriptors in terms of scalability and to make the result far more accurate than the proposed method.

REFERENCES

[1] Arun Hampapur, Ki-Ho Hyun and Ruud B. "Comparison of Sequence Matching Techniques for Video Copy Detection," Proc. SPIE, Storage and Retrieval for Media Databases, vol. 4676, pp. 194-201, 2002.

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

[3] Geert Willems, Tinne Tuytelaars and Luc Van G. "Spatio-Temporal Features for Robust Content-Based Video Copy Detection," Proc. ACM Int'1 Conf. Multimedia Information Retrieval (MIR), pp. 283-290, 2008.

[4] Heng Tao Shen, Jie Shao, Zi Huang and Xiao fang Z. "Effective and Efficient Query Processing for Video Subsequence identification," IEEE Trans. Knowledge and Data Eng., vol. 21, no. 3, 2009.

[5] Hong Liu, Hong Lu and Xiangyang Xue "A Segmentation and Graph-Based Video Sequence Matching Method for Video Copy Detection," IEEE Transaction on Knowledge and Data Engineering, vol. 25, 2013.

[6] Hong Liu, Hong Lu and Xiang yang X. "SVD- SIFT for Web near duplicate image detection," Proc. IEEE Int"1 Conf. Image Processing (ICIP '10), pp. 1445-1448, 2010.

[7] Julien Law-To, Li Chen and Alexis J. "Video Copy Detection: a Comparative Study," Proc. ACM Int'1 Conf. Retrieval, pp. 371-378, 2007.

[8] Li Chenal and Stentifordb. F.W.M. "Video Sequence Matching based on Temporal Ordinal Measurement," Pattern Recognition Letters, vol. 29, no.13, pp. 1824-1831, 2008.

[9] Mattijs Douze, Herve Jegou and Cordelia S. "An Image-Based Approach to Video Copy Detection with Spatio-Temporal Post-Filtering," IEEE Trans. Multimedia, vol. 12, no. 4, pp. 257-266, 2010.

[10] Onur Kucuktunc, Muhammet Bastan, Ugur Gudukbay and Ozgur U. "Video Copy Detection using multiple visual cues and MPEG-7 descriptors," J. Visual Comm. Image Representation, 2010.

International Journal of Innovative Research in Computer and Communication Engineering

BIOGRAPHY

Ms.Girija.K received the B.E degree in Computer Science from SKP Engineering College, Tiruvannamalai in 2012. She is currently doing her M.E in the same stream in Pavendar Bharathidasan College of Engineering and Technology, Tiruchirappalli.

Mr.Herman Jeeva S received the B.E degree in Computer Science and Engineering from St.Joseph's College of Engineering and Technology, Thanjavur in 2012. He is currently doing the M.E in the same stream in Pavendar Bharathidasan College of Engineering and Technology, Mathur, Tiruchirappalli.

Ms.Soniya.M received the B.E degree in Computer Science and Engineering from Anna University of Technology, Tiruchirappalli in 2012. She is currently doing her M.E in the same stream in Pavendar Bharathidasan College of Engineering and Technology, Tiruchirappalli.

Mr.Sabarinathan.P received his B.E degree in Computer Science from Annai Mathammal Sheela Engineering College, Namakkal in 2007 and received his M.E degree in the same stream in 2010 from Dhanalakshmi Srinivasan Engineering College, Perambalur. He is currently working as an Assistant Professor in Pavendar Bharathidasan College of Engineering and Technology, Tiruchirappalli and his area of interest includes MANET and networks.