International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 04 Issue: 11 | Nov -2017 www.irjet.net p-ISSN: 2395-0072

Real-Time Video Copy Detection in Big Data

Priyanka Prajapati[1], Meenal Mishra[2], Snehal Patil[3], Mamta Pawar[4]

[1,2,3,4] *B.E Computer Engineering, Terna Engineering College, Mumbai university, India*

Abstract - *With the evolution of new technology and facilities today more number of videos are uploaded all over the internet but it is not necessary that every video holds original contents. Video copy detection basically deals with finding similarities between original and duplicate video. Existing video copy detection algorithms have certain limitations such as it takes more searching time and sometimes gives inaccurate result. We have studied two video copy detection algorithms which will be implemented using Hadoop framework which is the method based on brightness sequence and TIRI-DCT. These are followed by a fast approximate search algorithm i.e. cluster-based similarity search. Existing exhaustive search method is time consuming process. All of these characteristics could just meet the requirements of real-time video copy detection technology.*

***Keywords*- Hadoop, Cluster-based similarity search, Inverted file-based similarity search, TIRI-DCT.**

1. INTRODUCTION

Due to rapid development of multimedia hardware and software technologies, the cost of image and video data collection, creation, and storage is becoming low. Every day lots of video data are generated and published. Among these huge volumes of videos, there exist large numbers of copies. According to statics 27 % redundant videos are duplicated on the most popular version of a video in the search results from video search engines. Therefore an effective and efficient method for video copy detection has become more and more important. Also users are often frustrated when they need to spend their valuable time to find videos of interest; they have to go through number of duplicated or nearly duplicated videos that are streamed over the internet before arriving at an interesting one. Because of these duplicated video lots of storage space and users time can be wasted.

Video copy is not a simple copy of the original video but a new video formed after it is attacked, which is visually not very different from the original one. The attack forms of video mainly include adding logo, improving contrast, reducing contrast, changing video size and so on. Video copy detection is to detect the video so as to determine whether it is the copy of another video or not. Following paper is a brief summary of the work carried out in this domain. Various video copy detection algorithms which include spatial domain and transform domain algorithms are reviewed here.

Video copy detection is divided into two parts. First, the features of the video are extracted and a hash library (Hadoop distributed platform is applied to calculate the hash values) of the videos is formed; secondly, the features of the querying video are extracted and a hash value of the querying video is formed and then it is compared with the hash values of the training videos, so that whether the query is a copy video or not can be determined [1].

2. LITERATURE SURVEY

In[4], we studied about watermarking. The limitations of watermarks are that if the original image is not watermarked, then it is not possible to know if other images are copied or not.

In[1], Each video frame has their spatial characteristics, and the combination of the video frames has time-domain characteristics. We have also studied about the fingerprint extraction techniques.

In[2],Size of online video databases can reach tens of millions of videos, which translates into a very large fingerprint database size. This means that even if the fingerprint of a query video can be extracted very quickly, searching the fingerprint database to find a match may take a long time. For online applications, however, a reliable match should be found in almost real-time. Therefore fingerprint matching forms a practical bottleneck for online fingerprinting systems.

In[6], A simple exhaustive search method which has a complexity of , where is the number of the fingerprints in the database. Search methods inverted file-Based Similarity Search is modified version of search algorithm that was proposed.

In[5], We learned about fingerprinting technology and application of video copy detection system in managing copyrighted content, video tracking, identifying. This paper has done considerations, such as robustness, compactness and discriminability for finger printing methods. It also discusses fingerprint matching technologies complexities.

3. CONTENT BASED COPY DETECTION

It is another method for identification of images and video clips. It is based on idea of "the media itself is the watermark," i.e., the media like video, audio, image contains

enough unique information for detecting copies. Signature is extracted from the trained media and stored in a database. The same procedure of signature extraction is done for query media and compared to the trained media signature which is already extracted and stored in a database to determine if the test stream contains a copy of the query media. The signatures which are extracted are called as "Fingerprint".

3.1 Fingerprint

Fingerprints summarize a video signal. These signature features vectors that uniquely characterize the specific signal. Video fingerprinting is a technique that can be used for content based copy detection. The major task is to detect whether a particular part of the video is based on the same original video in a database of reference video.

Temporal fingerprints-Initially, a video sequence is divided into shots. Then, the length of each shot is considered as a temporal signature, and the sequence of concatenated shot durations form the fingerprint of the video. This feature works well with long video sequences, but does not perform well for short video sequences since they do not contain enough information required.

Spatial fingerprints– This algorithm converts a video image into YUV colour space in which the luminance (Y) component is kept and the chrominance components (U, V) are discarded. Spatial fingerprints are unique value which derived from each frame or from a key frame.

Spatio-temporal fingerprints- One disadvantage of spatial fingerprints is that they are not able to capture the video's temporal data, which is an important differentiating factor[3].

3.2 Algorithm

We have studied the following algorithms of content-based copy detection.

a)TIRI-DCT Algorithm

Temporally informative representative images - discrete cosine transform (TIRI-DCT) is developed. TIRI-DCT is based on temporally informative representative images which contains spatial and temporal information of a video sequence.[2] This method calculates a weighted average of the frames to generate representative images. This sequence will carry the temporal as well as spatial information. The image is then divided into blocks. The first horizontal and the first vertical DCT coefficients (features) are then extracted from each block. The value of the feature from all the blocks is concatenated to form the feature vector. Each feature is then compared to a threshold (which is the median value of the feature vector) and then a binary fingerprint is generated.

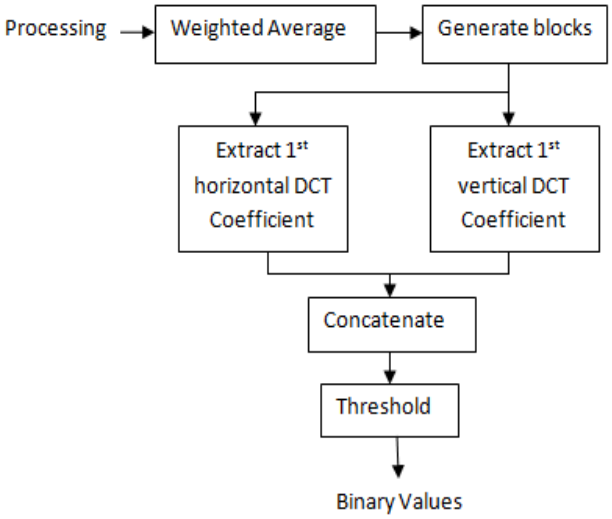

Figure 1: Schematic of the TIRI-DCT algorithm.

In order to determine whether a query video is an attacked version of a video in a database or not, it's fingerprint is first extracted. The fingerprint database (previously created from the videos in the video database) is then searched for the closest fingerprint to the extracted query fingerprint.

b) Brightness sequence Algorithm

The brightness of each frame is calculated of the video whose originality is to be found. This brightness values are compared with the values of the brightness of the original video at the same instant of time. If the brightness of both the videos is almost same, then it is said to be copied.

Step 1: Video hash extraction

For a video sequence X = x1,x2,x3,……xn, N is the frame number of the videos. The hash sequence corresponding to Video X is H = [f(x1),f(x2),…..f(xn)], f(x) = [a1,a2,a3….am] is the hash extracting function, M is the number of hash bits extracted from a single frame. Hash extraction algorithm is:

1: The video frame xi is divided into w* h areas
2: Calculate the average brightness of each block
3: Sort the brightness of each block
4:Now i= i +1, repeat step1, end the process until i= N.

Step 2 : Video hash matching

For video sequences X = x1,x2,x3,……xn, and Y = y1,y2,y3,……yn,, the hash sequences corresponding to videos X and Y are Hx = [f(x1),f(x2),…..f(xn) and Hy = [f(y1),f(y2),…..f(yn) respectively, wherein f(x) = [a1,a2,a3….am] and f(y) = [b1,b2,b3….bm] .

The distance between the two video sequences is calculated ie D(X,Y).When the distance D(X,Y) =T , the two videos are regarded as the same video. T is the predetermined threshold.

4. SEARCHING TECHNIQUE

4.1 Inverted-File-Based Similarity Search

Step 1: The binary unique values i.e. fingerprint are divided into n words each of same bits.

Step 2: The horizontal length of the table is used to represent the word position.

Step 3: The vertical length of the table is used to represent the possible values of words.

Step 4: Index is added to each word of the fingerprints for entry in column corresponding to the value assigned to the word.

Step 5: The calculation of hamming distance is done between the fingerprints in the database and the query fingerprints.

Step 6: If the calculated distance is less than the threshold value the query video gets matched and is declared as matching.

Step 7: Otherwise, it will be declared as videos not matching.

4.2 Cluster-based similarity Search

- This is another algorithm for binary fingerprints.
- The idea behind this is to make use of clustering to decrease the number of queries that are analysed within the database.
- If the fingerprints are assigned to only one cluster, then the fingerprints in the database will be clustered into non-overlapping groups.
- To perform the same, for each cluster a centroid is chosen and named as cluster head. The cluster will be assigned a fingerprint if it is closest to this cluster's head.
- The cluster head closest to the query is found to determine if the query fingerprint matches a fingerprint in the database.
- The fingerprints which belong to this cluster are searched to find the match i.e., the one which has the minimum Hamming distance (of less than a certain threshold) from the query. If match is not found, then the cluster which is the next closest to the query is examined [2].

5. ADVANTAGES

a. The performance on detecting copies from large data set is satisfactory and Hadoop platform can significantly improve the efficiency of video copy detection.
b. It has high fault tolerance, high throughput, easy scalability and etc.
c. The measurement of copy detections performance System making result is faster than existing system.
d. The video hashing algorithm methods has strong robustness, high distinction, high compactness and low complexity.

6. DISADVANTAGES

a. In existing system the fingerprint which is extracted using TIRI-DCT is not robust against the content changing attacks such as changing the background of the video or replacing picture in picture and the performance of the system is low in the presence of some other attacks, such as cropping, and logo insertion.
b. The security of the fingerprint can be achieved only with the fingerprints of smaller length. Larger fingerprints results in a decrease in detection speed as they require more computation in calculating the hamming distance between the fingerprints.

7. CONCLUSION

In this paper, optimization of both the techniques i.e. TIRI-DCT and Brightness Sequence Algorithm, their correctness, speed and efficiency are analysed. The efficiency of Hadoop platform is also focused which has high processing speed and is optimum for storing and retrieval of data.

REFERENCES

[1] Jing Li, XuquanLian, Qiang Wu and Jiande Sun "Realtime Video Copy Detection Based on Hadoop," Sixth International Conference on Information Science and Technology Dalian, China; May 6-8, 2016.

[2] Vaishali V. Sarbhukan, Prof. V.B.Gaikwad ,"Video FingerprintingExtraction Using TIRI-DCT" , International Journal of Engineering Researchand Development 2013

[3] AbrumJeysudha,SamitShivadekar, "Real Time Video Copy Detection usingHadoop",International Journal of Computer Application 2017.

[4] J. Law-To, L. Chen, A. Joly, I. Laptev, O. Buisson, V. Gouet-Brunet, N. Boujemaa, and F. Stentiford, Video copy detection: A comparative study, in Proc. Conf. Image Video Retrieval (CIVR), 2007.

[5] Jian Lu, "Video fingerprinting for copy identification: from research to industry applications", Proceedings of SPIE - Media Forensics and Security XI, Vol. 7254, January 2009.

[6] Ms. Laxmi Gupta, Prof. M.B Limkar, Prof. Sanjay. M Hundiwale, "Search Methods for Fast Matching of Video Fingerprints within a Large Database",by SSRG International Journal of Electronics and Communication Engineering (SSRG-IJECE) – volume1 issue 3 May 2014