# REAL TIME COPIED VIDEO DETECTION FOR PUBLIC GREIVANCE WEBSITE

## POONAM BHAGADE[1], MADHURI KAMBLE[2], SNEHA CHINDARKAR[3]

*[1, 2, 3] B.E. Computer Engineering, Shivajirao S. Jondhale College of Engineering, Mumbai, India*

-------------------------------------------------------------------------***----------------------------------------------------------------------

**Abstract -** *With the development of multimedia technology and Internet, the amount of videos in the Internet is increasing quickly. Among the large amount of videos in the Internet, a considerable number of them are copies of original videos, which are simply revised versions of the original ones. The purpose of video copy detection technology is to detect copy videos, which has important applications in video tracking, video content retrieval, video copyright protection and other aspects. Hadoop is a distributed computing platform which is designed for deployment in inexpensive hardware and suitable for those applications with a large data set. All of these characteristics could just meet the requirements of real-time video copy detection technology. In this paper, an attempt is done to develop a real-time video copy detection system based on Hadoop platform for Public Grievances Website, and two video copy detection algorithms are implemented on Hadoop platform, which are the method based on brightness sequence and the method based on TIRI-DCT respectively, and their performances are compared. Experiments show that the use of Hadoop platform can significantly improve the efficiency of video copy detection, which has important practical significance for video tracking and real-time video content retrieval application.*

*Key Words***:  Hadoop, TIRI-DCT , Website, Videos, Public Grievances, Copy Detection.**

## 1. INTRODUCTION

Nowadays, the Internet are playing more and more important role in peoples life. The abundant resources in the Internet are expressed in many forms such as text, picture, audio, and video and so on. Among all of these kinds of media, video is loved by people because of its merits of clear expression and interest. However, there are many videos with similar content, which increases the burden on the network. Video copy detection is to detect the similarity of two video contents, i.e., to judge whether the two videos are from the same content by calculating the hash values of them.

A grievance is a formal complaint that is raised by public towards government or private organization in the workplace. There are many reasons as to why a grievance can be raised, and also many ways to go about dealing with such a scenario. The government and the private organization uses an internet based approaches such as Public grievance remedy portal - India, TripAdvisor, and Act Please. Just like above private organizations we have created

our website for video copy detection of public grievance (e.g. government jobs).

Hadoop platform is a commonly used distributed platform, which uses the Map Reduce programming model so as to greatly simplify distributed programming and uses the HDFS (Hadoop Distributed File System) distributed file storage system so as to allow more reliable file storage.

In this paper, the Hadoop distributed platform is applied to calculate the hash values of a large number of videos and then match them. Two video copy detection algorithms which are based on brightness sequence and TIRI-DCT respectively are implemented on Hadoop platform, and their performances are compared with in experiments.

## 1.1 Introduction to Hadoop Platform

Hadoop was developed by the Apache Foundation. The developers of Hadoop do not need to understand the underlying implementation of a distributed architecture and only need to follow the MapReduce programming model to develop the distributed applications. Hadoop has advantages of high fault tolerance, high throughput, easy scalability and etc. Hadoop mainly includes MapReduce, HDFS, Pig, Hive, HBase and Zookeeper technology.

In a Map Reduce process Map function processes the input fragmentation and outputs the processing result to the Reduce function; Reduce function combines the outputs of the Map function. The Map function takes a key pair (Key-Value) as input and generates a set of intermediate key pairs, the Map Reduce framework passes the intermediate key pairs generated by Map function in every Task Tracker to the Reduce function, and then the Reduce function combines these key pairs to produce a smaller key pair.

HDFS is the distributed file storage system of Hadoop. It consists of a Name Node and Data Node. Pig is a high-level query language built on Map Reduce, which can simplify the development process of Map Reduce. Hive is a data warehouse tools on Hadoop, which can provide SQL queries. HBase is a distributed database, which is based on the column storage model. Zookeeper provides coordination services in distributed systems on Hadoop.

## 1.2 Introduction to Existing System

All existing solutions to analyses video streams are based on real time event detection using computer vision algorithms, which help in detection of abnormal events, which in turn trigger alarms for a variety of user defined events. The core of such a system is the increasingly intelligent and robust video analytics that are capable of analyzing the videos from low-level image appearance and feature extraction to midlevel event detection to high-level reasoning and scene understanding. But in order to analyses historic data, which may allow us to gather deeper insights into data.

## 2. PROPOSED SYSTEM

We propose a video copy detection using method based on Brightness sequence and the method based on TIRI-DCT algorithm. In this proposed system, video copy detection is divided into two parts. First, the features of the training video are extracted and a hash library of the training 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.
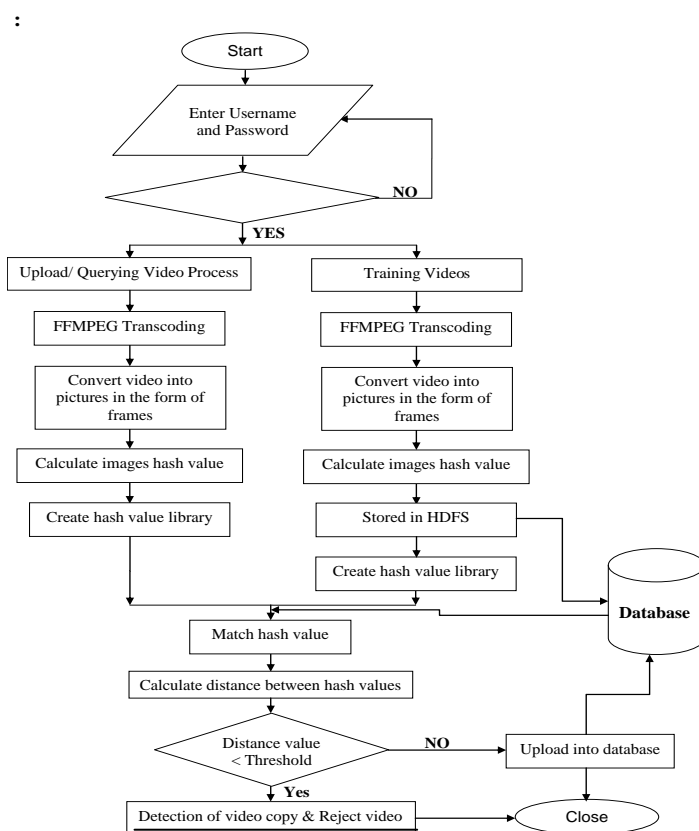
## 3. SYSTEM ARCHITECTURE

:



**Fig -1**: Flowchart

The admin can login into the system with given username and password stored in the database. If the password matches, the admin is authorized and given access to the system. The original videos, which are also known as training videos, are transcoded using FFMPEG. Since Hadoop doesn't support FLV format videos, it has to be converted to different format if in case the admin upload the videos in FLV format.

## 3.1 TIRI-DCT Algorithm

The first method of finding the copied videos is by using TIRI-DCT algorithm. In this approach, at each instant of time of a particular interval, a frame is drawn. The features of that particular frame at that instant of time are extracted. The extracted characteristics are then compared with the characteristics of the original videos present in the database.

If the features of both the videos are almost same, then the video is said to be copied. For example, suppose a particular video has four frames. If the feature of one frame matches with the original video, it is said to be 25 % copied video. If two frame matches, it is said to have 50% originality and so on.

## 3.2 Brightness Sequence Algorithm

The second method is Brightness Sequence Algorithm. In this approach, 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.

$$R_{x,y} = \Sigma J_{k-1}\ 0.65^k * l_{k,x,y} \ldots\ldots [1]$$

Where $R_{x,y}$ = Frame

$l_{k,x,y}$ = Brightness value of the kth frame at the

location (x, y)

This algorithm is implemented on the frames extracted from the video in hand. The hash values of the same is calculated and then the distance between the hash values of querying video as well as the training video is determined which is then compared against a predefined threshold value. If the distance is greater than the predefined threshold value, then the two videos are considered as different videos else they are regarded as same video, hence the copied one

## 4. IMPLEMENTATION

In implementation phase, similarities and dissimilarities between the videos are found out using Hadoop platform and by making use of two phases, i.e. Map phase and Reduce phase. The two distinct algorithms which are being used are TIRI-DCT and Brightness Sequence Algorithm. The main aim of these algorithms is to convert the video into multiple

frames, extracting their features and then finally comparing both querying as well as training video. The training video is already present in the HBase, while the querying video is taken as input from the user.

## 5. FUTURE WORK

Further improving the scalability of the system. We will investigate other local visual features, including SURF, which take less time to compute. Global visual features, including color histogram, edge and texture information, may also be used as a pre-processing step to speed up the local visual feature search.

Incorporating the audio information. Videos normally contain both audio and video streams. The audio information can be used to further improve the detection accuracy and reduce the computational complexity. We are planning to use a set of audio features, including Mel-frequency cepstral coefficients (MFCC), and evaluate the performance of pattern recognition schemes, such as, dynamic programming and hidden Markov models. It will be interesting to evaluate the fusion of audio and visual detection results as well.

Applying the developed copy detection algorithm in real world applications. Possible applications and services include large scale content based image and video search, video copy detection for P2P streaming network, etc.

## 6. CONCLUSION

In this project we have studied about video copy detection based on Hadoop. The algorithms are implemented on Hadoop distributed computing platform and the efficiencies are compared in different video amounts and different map amounts. The further work is to research more video copy detection algorithms and make comparisons between them; optimize the processing of the algorithms in Hadoop platform; make good use of HBase distributed database to fasten the extraction and retrieval of video hash; extend the algorithms to the copy detection for image and text

## 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] K. Iwamoto, E. Kasutani, and A. Yamada. "Image signature robust to caption superimposition for video sequence identification," Proceeding of the 2006 IEEE International Conference on Image Processing, pp. 3185-3188, 2006.

[6] L. Chen, and F. W. M. Stentiford. "Video sequence matching based on temporal ordinal measurement," Pattern Recognition Letters, vol. 29, no. 13, pp. 1824-1831, 2008.