Real Time Video Copy Detection on bigdata with no Hadoop

[1]Prof. Sumeet Pate, [2]Yogesh Ramesh Kolte, [3]Pranav Mohan Kadam, [4]Rohit Dileep Oza,

[1]Asst.Professor,[2,3,4]UG Student,[1,2,3,4]Computer Engg.Dept, Shivajirao S. Jondhle College of Engineering & Technology, Asangaon, Maharashtra, India.

[1]sumeetpate09@gmail.com, [2]y.kolte27@gmail.com, [3]pranav.kadam03@gmail.com, [4]oza.rohit78@gmail.com.

Abstract--Bags of videos are getting uploaded to the internet from time to time and are aggregate every day. Out of these videos, ample numbers of videos are actionable copies or some videos are manipulated versions of absolute media. Because of rising enthusiasm for videos, there are different locales which furnish with various types of videos however it isn't vital that each video hold unique content. Video Copy Detection basically deals with award out similarities amid the agreeable of two accustomed videos.

Keywords— MD5, querying video, training video, hash value.

1. INTRODUCTION

Nowadays, media has moved toward becoming an important allotment of everyone's circadian life. The aim of video copy detection is to choose whether an inquiry video section is a duplicate of a video from the video dataset. Due to arising absorption in videos, there are assorted sites which accommodate with altered kinds of videos but it is not all-important that every video authority aboriginal content. Hence, Video Copy Detection action comes into the record to differentiate amid aboriginal and alike videos. This can be fitted by calculating the hash values of the content of the videos by using appropriate algorithms.

This technology has editing and duplication of video data that will cause to violation of digital rights. Due to billions of videos present on the web, it is not accessible to run the video copy detection process on a single machine approach as it is a hectic process. Due to a bulk of calculations present in this process, a distributed computing approach will fetch efficient result as compared to that of single machine approach. [1,2] To abundance and action an ample bulk of data, Hadoop which is accessible antecedent and Java programming accent plays an major role as it works on the appropriate environment.

Aims & Objectives

1. The aim of the project is to provide security to multimedia content and provide platform to safeguard copyright of the digital media.
2. As the digital videos can be duplicated and modified easily, ensuring the copyright of advanced media must be our first priority.
3. Protection of content from fraudulent alterations.
4. Protection against copyright violations.

III. LITERATURE SURVEY

[1] An Efficient Content and Segmentation Based Video Copy Detection

This paper first examines the features used for video copy detection. Through the analysis done, they use common feature derived by SIFT-SURF matching algorithms to describe video frames. Then, they use a dual-threshold adjustment to annihilate bombastic video frames and use the SIFT-SURF analogous algorithm to compute the affinity of two SIFT-SURF affection point sets. Besides, for video succession coordinating, they utilized a chart-based video grouping technique for usage. [1] Thus, skillfully changes the sequence corresponding result to a matching output graph.

[2] Robust Video Fingerprinting for Content-Based Video Identification

This paper is suited to use a video fingerprinting adjustment based on the Centre of mass of acclivity orientations This method is not only identically independent but also power full against accepted video processing accomplish including lossy contraction, resizing, color, gamma etc. The problem of reliable matching is proceeding by assuming the fingerprints an ability of an anchored ergodic process. [4] The coordinating edge is hypothetically inferred for a given false alert rate utilizing the accepted stochastic model, and its legitimacy is tentatively checked. This paper tends to plan a safe video fingerprinting strategy powerful against general geometric changes.

[3] Enhancing Security in a Video Copy Detection System Using Content Based Fingerprinting

This paper explains fingerprint abstraction algorithm followed by an almost seek method. This Interest Point Matching Algorithm extracts power full, discriminate and compact fingermark from videos in a quick and reliable fashion. [3] This fingerprint extracted using this algorithm maintains an acceptable achievement for attacks such as commotion expansion, changes in brightness or contrast rotation, time move, and changes in the foundation. The algorithm will reduce the searching time so it improves to faster searching methods.

IV. EXISTING SYSTEM

The current arrangement of video duplicate location needs more data and the more profound examination of the video. The important approaches in existing system which is used is watermarking technique. Thus, undetectable signals are included into the videos. Amid the recognition, the videos are changed over into different pictures which have watermark on it, which helps in the identification procedure. The constraint of this procedure is if the first video isn't watermarked, it is difficult to know whether the video is duplicated or not. Unique features are removed from the video and are coordinated with the current features show in the database. If they are same, the video is said to be duplicated.

V. PROBLEM STATEMENT

Watermarking is the traditional strategies dealt with down for quite a while to represent the responsibility for specific video by methods for embeddings the logo or some other advanced mark of the owner. The problem is that data is either unmistakable to the human eye or even can be covered up. This watermarking data can be expelled effectively by anybody by making any little change to the watermarked video which incorporates re-encoding and change of bit rate. It is critical to know that watermarking isn't bound to recover a question video cut.

VI. PROPOSED SYSTEM

Video copy detection is a commutual access to watermarking technique. Since watermarks must be reason that the original content before duplicates are made, it can't be connected to content which is as of now available for use. The proposed system uses MD5 algorithm. MD5 is a cryptographic hash function whose fundamental intention is to check that a document has been unaltered. Instead of acknowledging that the two sets of videos are identical to raw data, MD5 does this by bearing a assortment amount of both sets and again comparing to verify that they are same.MD5 processes a variable-length message into a fixed-length output of 128 bits. The algorithm takes as input a message of arbitrary length and produces as output a 128-bit message digest. The input is processed in 512-bit blocks.

Step 1: Appending padding bits
Step 2: Append length
Step 3: Initialize MD buffer.
Step 4: Process message in 512-bit (16-word) blocks.
Step 5: After all, 512-bit blocks have been
processed, the output is 32-digit hexadecimal number.

```
typedef struct {
  char data[64];
  int datalen;
  int bitlen[2];
  int state[4];
} MD5_CTX;
```

Step 1: MD5 Initialization

```
md5_init(MD5_CTX *ctx)
{
  ctx->datalen = 0;
  ctx->bitlen[0] = 0;
  ctx->bitlen[1] = 0;
  ctx->state[0] = 0x67452301;
  ctx->state[1] = 0xEFCDAB89;
  ctx->state[2] = 0x98BADCFE;
  ctx->state[3] = 0x10325476;
}
```

Step 2: MD5 message-digest operation, processing another message block, and updating the context.

```
void MD5Update (context, input, inputLen)
MD5_CTX *context;
unsigned char *input;
unsigned int inputLen;
unsigned int i, index, partLen;
index = (unsigned int)((context->count[0] >> 3) & 0x3F);
if((context->count[0] += ((INT4)inputLen << 3))<
((INT4)inputLen <<3))
context->count[1]++;
context->count[1] += ((INT4)inputLen >> 29);
partLen = 64 - index;
if (inputLen >= partLen) {
MD5_memcpy((POINTER)&context->buffer[index],
(POINTER)input, partLen);
```

Step 3: Transforms the block and saves the state to provide message digest.

```
MD5Transform(context->state, context->buffer);
for (i = partLen; i + 63 < inputLen; i += 64){
MD5Transform (context->state, &input[i]);
index = 0;
}
else
i = 0;
MD5_memcpy((POINTER)&context->buffer[index],
(POINTER)&input[i],inputLen-i);
}
```

Step 4: Ends an MD5 message-digest operation by padding the last block till 448 bits then appending data-length in bits to the last 64 bits to get final message of 128 bits.

```
md5_final(MD5_CTX *ctx, char hash[])
{
  int i;
  i = ctx->datalen;
  if (ctx->datalen < 56) {
    ctx->data[i++] = 0x80;
    while (i < 56)
      ctx->data[i++] = 0x00;
  }
  else if (ctx->datalen >= 56) {
```



```
    ctx->data[i++] = 0x80;
    while (i < 64)
    ctx->data[i++] = 0x00;
    md5_transform(ctx,ctx->data);
    memset(ctx->data,0,56);
  }
}
```

VII. MATHEMATICAL MODEL

A message digest is a cryptographic hash function containing a string of digits. Message digests are intended to secure bit of information or media to recognize changes and modifications to any piece of a message. Message digest hash numbers represent specific files containing the protected works. One message digest is assigned to particular data content. It can reference a change made intentionally or coincidentally, yet it prompts the owner to recognize the alteration and also the individual rolling out the improvement. Message digests are algorithmic numbers. This term is also known as a hash value or checksum.

1]Create MessageDigest object for MD5.
2]Update input string in message digest.
3]Converts message digest value in base 16.
Video

$$\text{Copy\%} = \frac{(\text{Total hash value match in videos}) \times 100}{\text{original video size}}$$

VIII. SYSTEM ARCHITECTURE

The modules in this paper are utilized for a successful investigation of content-based video duplicate discovery is made out of three sections:

A. Pre-Processing:
In this first module, they select the video from the video database and then split the particular video into frames. After separating the video into frames, modify the frames into Luminance and Chrominance Valued (LUV) frames.

B. Short Segment Formation:
In the second module, after pre-preparing the video frames are isolated into covering sections of settled length, each containing j frames. At that point, algorithm is implemented to these section and hash value is generated for corresponding frames.

C. Hash Value Comparison:
In the third module, Comparison of hash values for querying and training videos is done. As per matching result of hash values for both videos obtained, corresponding result is displayed.

IX. ADVANTAGES

1.It can consequently discover ideal succession coordinating after effect of videos.
2. It can naturally evacuate the noise caused by visual component coordinating.
3. It is adaptive to video anatomy amount change.
4. It can distinguish numerous duplicates existed in the recognized video.
5. This strategy considers the spatiotemporal normal for the video succession
6. This strategy can locate the best coordinating the arrangement in numerous match results comes about which successfully avoids false "high similarity" commotion and repay the constrained depiction of picture visual highlights.

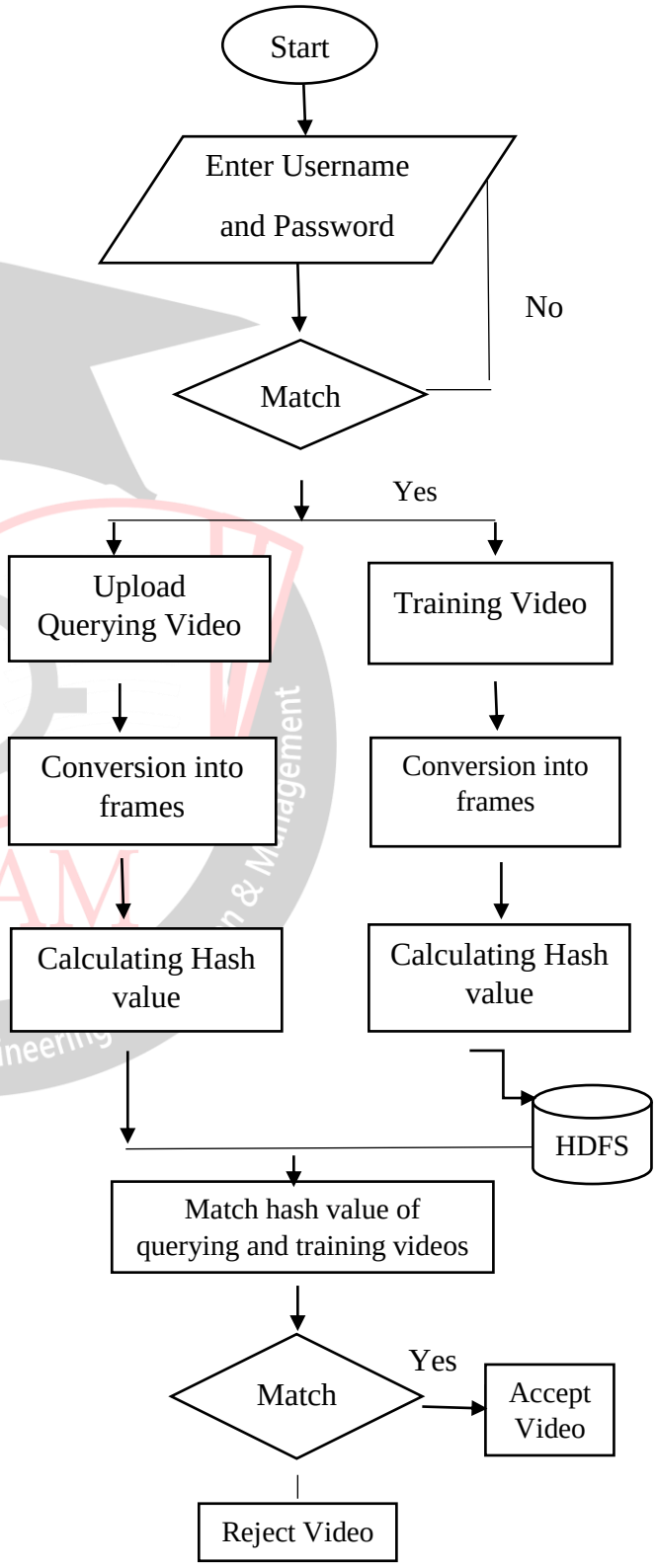

Fig 1: System Architecture

.

TableNo.1: Comparative Study

Sr. No	Paper Title	Author's Name	Problem	Solution
1.	A Robust Video Copy Detection System using Tiri-Dct and Dwt Fingerprints	Devi S, Vishwanath and S. Muthu Perumal Pillai	Performance is low for rotation and shift attacks	Combination of feature based matching and inverted index files
2.	An Efficient Content and Segmentation Based Video Copy Detection	N. Kalaiselvi, K. Priyadharsan	Searching longest path in matching result graph	Dynamic programming method
3.	Robust Video Fingerprinting for Content-Based Video Identification	Sunil Lee, Member, IEEE, and Chang D. Yoo, Member	Single fingerprint with low dimension is not suitable for reliable matching	Fingerprint sequence is generated by concatenating the fingerprints extracted from frames
4.	Robust Video Signature based on ordinal measure	Xian-Sheng Huu, Xian Chen and Hong-Jiung Zhnng	Not able to detect changes in rotation and scale invariant in video	Combination of SIFT and SURF features must be used
5.	Enhancing Security in a Video Copy Detection System Using Content Based Fingerprinting	E. Meenachi, G. Selva Vinayagam, C. Vinothini	Time complexity is more due to large database searching	Locality Secure Hashing

X. RESULT & DISCUSSION

The user enters username and password considering various parameters listed in the front end. If password matches, the admin is authorized and get access to the system. First, upload original video also known as training video then compare it with querying video that is taken as input from user. If the original video and querying video both contain some hash values then by matching hash values of both videos, the result is displayed.

Fig 2: Matching Result Page

XI. CONCLUSION

We have to tried to implement "Real Time Video Copy Detection on big data with no Hadoop" International Conference of Computer Applications, Volume 162 – No 9, March 2017 paper and after implementation we got the conclusion as: This paper explains similarities and dissimilarities between the videos. The capital aim of these process is to convert the video into assorted frames, calculating their hash values and again assuredly comparing both querying as able-bodied as a training video. The preparation video is as of now exhibit in the HBase, while the questioning video is taken as contribution from the client. This algorithm has a top accurate absolute amount for these ambits of attacks while it is actual authentic with a low apocryphal anxiety rate. It is as well computed decidedly fast compared to abounding absolute video hashing algorithms. Detection of the affected video based on its capacity includes use of circuitous algorithm which is disadvantage of such address because they need added time to execute. They charge to plan on the segments and again called key frames are acclimated to abstract the appearance value. There is charge to advance the new technologies which may detects affected video inn beneath time and provides bigger protection.

REFERENCES

[1] X.-S. Hua, X. Chen, and H.-J. Zhang. "Robust video signature based on ordinal measure" In International Conference on Image Processing, 2004.

[2] S. Lee and C. Yoo, "Robust video fingerprinting for content-based video identification," IEEE Trans. Circuits Syst. Video Technol., vol. 18, no. 7, pp. 983–988, Jul. 2008.

[3] G. Willems, T. Tuytelaars, and L.V. Gool, "Spatio-Temporal Features for Robust Content-Based Video Copy Detection," Proc.ACM Int'l Conf. Multimedia Information Retrieval (MIR), pp. 283-290, 2008.

[4] M. Malekesmaeili, M. Fatourechi, and R. K. Ward, "Video copy detection using temporally informative representative images," in Proc. Int.Conf. Machine Learning and Applications, pp. 69–74, Dec. 2009.

[5] Anindya Sarkar, Vishwakarma Singh, Pratim Ghosh, Bangalore S. Manjunath, and Ambuj Singh, "Efficient and Robust Detection of Duplicate Videos in a Large Database," IEEE Trans. on Circuits and Systems for Video Technology, vol. 20, no. 6, (2010)

[6] E. Meenachi, G. Selva Vinayagam, C. Vinothini "Enhancing Security in a Video Copy Detection System Using Content Based Fingerprinting," IEEE Vol.2, Issue.4, pp-1747-1753, July-Aug. 2012.

[7] M. Ramya, R. Kanthvel "Efficient and Scalable Content- Based Video Copy Detection System" International Conference on Recent Trends in Computational Methods, Communication and Controls, IEEE 2012.

[8] Vaishali V. Sarbhukan, Prof. V. B. Gaikwad "TIRI – DCT Based Video Copy Detection System," IEEE, Vol. 2, Issue.6, pp.1544-1548, November-December 2012.

[9] N. Kalaiselvi, K. Priyadharsan, "A Efficient Content and Segmentation-Based Video Sequence Matching Method for Video Copy Detection," (Volume-III, Issue-XI) November-2014.

[10] Avinash H. Kakade, Prof. JadhavMeenal "Survey on Graph-Based Video Sequence Matching &BoFs method for Video Copy Detection," IEEE, Volume 2, Issue 6, October-November, 2014.

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

