Improving Content Based Video Retrieval Performance by Using Hadoop-MapReduce Model

El Mehdi Saoudi, Abderrahmane Adoui El Ouadrhiri, Othman El Warrak, Said Jai Andaloussi, Abderrahim Sekkaki
Hassan II University, Casablanca, Morocco
elmehdi.saoudi@gmail.com, {a.adouielouadrhiri-etu , othman.elwarrak-etu , said.jaiandaloussi, abderrahim.sekkaki}@etude.univcasa.ma

***Abstract*—In this paper, we present a distributed Content-Based Video Retrieval (CBVR) system based on MapReduce programming model. A CBVR system called Bounded Coordinate of Motion Histogram (BCMH) has been implemented as case study by using Hadoop framework. Our work consists of proposing a distributed model to extract videos signatures and compute similarity with the BCMH system based on a set of Mapreduce jobs assigned to multiple nodes of the Hadoop cluster in order to reduce computation time of training process. The proposed approach is tested on HOLLYWOOD2 dataset and the obtained results demonstrate efficiency of the proposed approach.**

I. Introduction

Nowadays, with the quick development of the social media, an impressive amount of video data are collected and generated, 300 hours of video are uploaded to YouTube every 60 seconds [1] and 10 billion videos are watched on Snapchat per day [2]. With this revolution, the volume of video data to be processed increases significantly as well as the computational time. The main objective of Content-Based Video Retrieval (CBVR) systems is to retrieve the videos similar to a query video from a video database using a set of criteria. CBVR systems are now used in different applications, especially CCTV systems and medical field. The literature proposes several CBVRS models, Gwenole Quellec et al. [3] presented a CBVR system for Real-Time Retrieval of Similar Videos with Application to Computer-Aided Retinal Surgery, to characterize video sub-sequences they extracted texture, color features and motion features, to compare videos a partial lowlevel distance is calculated. Snehal Anwekar, Isha Walimbe, Priyanka Suryagan, Alisha Gujarathi and Kalpana Thakre [4] presented a Content Based Video Surveillance System for crime prevention based on moving object detection in real time, they used blob detection methods for finding areas of interest and Centroid Variance method for motion detection. A CBVR system based on Bounded Coordinate of Motion Histogram (BCMH) is presented by El Ouadrhiri et al. [5] to characterize video sub-sequences, they use vector motions and residual data as signatures, and use the Bounded Coordinate System to calculate similarity between them, according to the results, the similarity measurement accuracy as well as the response time are good, but we noticed that once the amount of video data increases, the framework shows signs of weakness in time computation. This paper presents a system architecture for handling Big video data in CBVR systems, by using a distributed computing platform, with Hadoop [6] [7] system based on MapReduce framework. We also chose El Ouadrhiri et al. CBVR system as a use case for our approach. Therefore, optimizing the BCMH computation time is the challenge in the current scenario.

The rest of this paper is organized as follows. In section II we introduce the basic concepts and related work in CBVR systems using the Hadoop distributed platform. Section III presents our use case framework (BCMH). In the next section, we define our approach. section V describes the system architecture and his implementation. Section VI presents the experimental results of the solutions and analysis. Finally Section VII concludes this paper.

II. Related works

One of the effective frameworks which is used in handling Big Data is Apache Hadoop. It is an opensource framework invented by Doug Cutting and Mike Cafarella in 2005. That allows for the distributed processing of large data sets through clusters of computers using simple programming models [8]. It offers reliable, scalable and distributed computing. The most core technology of Hadoop is HDFS, YARN and MapReduce.

Hadoop Distributed File System (HDFS): Is a distributed file system for mass data and large files, that provides high-throughput access to application data, It achieves reliability by replicating the data across multiple hosts. [9]

Hadoop YARN: Is a framework providing resource management, job scheduling and a central platform to deliver consistent operations, security, and data governance tools across Hadoop clusters. [10]

Hadoop MapReduce: Is the programming model of Hadoop which includes Map function and Reduce function for parallel processing of large data sets [11] . Fig. 2 shows the complete architecture of Yarn, consists of a master daemon known as "Resource Manager", slave daemon called node manager (one per slave node) and Application Master (one per application).

Some CBVR researchers solve the computation time problem by using Hadoop distributed system. A Surgical Video Analysis method using Hadoop distributed system to detect instruments used in surgical was proposed in [12]. Moyun Li,

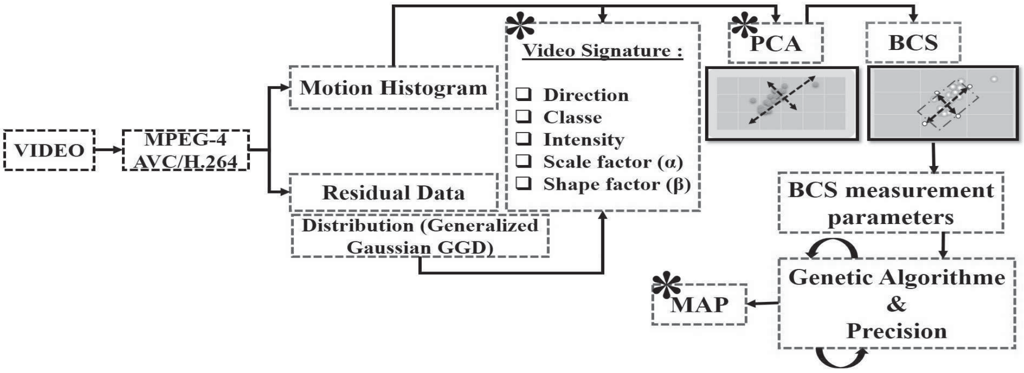

Fig. 1. Bounded Coordinate of Motion Histogram (BCMH) process

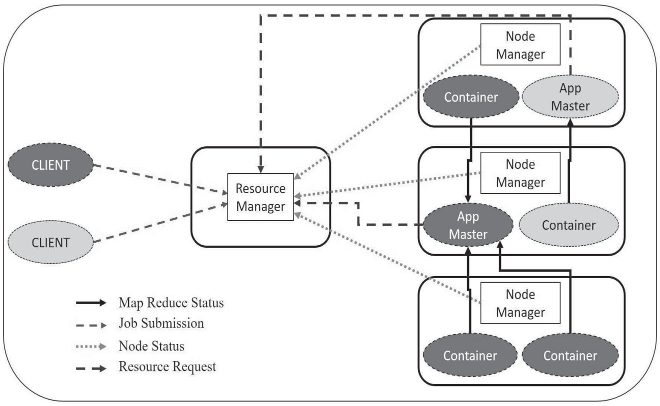

Fig. 2. Apache Hadoop Yarn Architecture

Cheng Yang and Jiayin Tian [13] employ a distributed system to improve the video encryption speed and optimize video encryption strategies. Also Tan Hanlin and Chen Lidong [14] proposed an approach for fast and parallel video processing on Hadoop distributed system, as case studies, face detection and motion detection and tracking algorithms have been implemented on clusters. Moula Husain et al. [15] proposed a distributed system for counting occurrences of each textual word from video frames. M. Mazhar Rathore, Hojae Son, Awais Ahmad and Anand Paul [20] proposed an approach for Real-time video processing for traffic control in smart city using Hadoop ecosystem with GPUs. JAI-ANDALOUSSI Said et al. [21] proposed a Medical Content Based Image Retrieval technique by Using the HADOOP Framework.

III. CBVR Based on Bounded Coordinate of Motion Histogram

The following Section summarizes the major components of the BCMH framework developed by El Oudrhiri et al., that will be used as use case in the rest of the paper.

To compute similarity between video sub-sequences, they proposed a Content-Based Video Retrieval (CBVR) solution called Bounded Coordinate of Motion Histogram (BCMH). The concept is to create signatures based on vector motions and residual data to characterize videos, and use the Bounded Coordinate System to measure similarity between them. Fig. 1 shows the complete BCMH process. Accordingly, the video signature becomes represented by two categories, motion vector histogram and the residual error, which represent the difference between predicted and original frames.

A. Motion histogram classification

The first part of video signature is represented by: 1) Direction: highest amount of vector μ on histogram. 2) Class: number of vector μ, dominant direction. 3) Intensity: the median of a total motion vectors of the dominant class is calculated with the equation (1) [5].

$$Intensity_{\mu} = \frac{1}{D}\sum_{i=1}^{D}|\mu|\,;(D:Direction) \tag{1}$$

The equation (2) calculate the direction of motion vector μ(x,y).

$$\Omega(\mu) = \begin{Bmatrix} \arccos\frac{\text{x}}{|\mu|} & ,y \geq 0 \\ 2\pi - \arccos\frac{\text{x}}{|\mu|} & ,y < 0 \end{Bmatrix} \tag{2}$$

This equation is available if μ(x,y) $\neq$ (0,0) with length $|\mu|$, thus, motion direction interval is contained M=13 bins of directions. 12 directions and a separate bin M=0 for zero-length motion vectors (Fig. 3) [16] , the motion histogram is calculated with the equation (3).

$$Histogram(\mu) = \begin{cases} 0 & ,\mu = (0,0) \\ 1 + ([\Omega(\mu)\frac{M}{2\Pi} + \frac{1}{2}]modM) & ,otherwise \end{cases} \tag{3}$$

B. Motion compensation of residual information

To get the second part of signature, which represents the residual information, they used the generalized Gaussian distribution methode (GGD) [17], defined by (4).

$$P(x,\alpha,\beta) = \frac{\beta}{2\alpha\Gamma(\frac{1}{\beta})}\,e^{-(\frac{|x|}{\alpha})\beta} \tag{4}$$

The gamma function is $\Gamma(x) = \int_0^{\alpha} e^{-t}t^{x-t}dt; x > 0$, where:

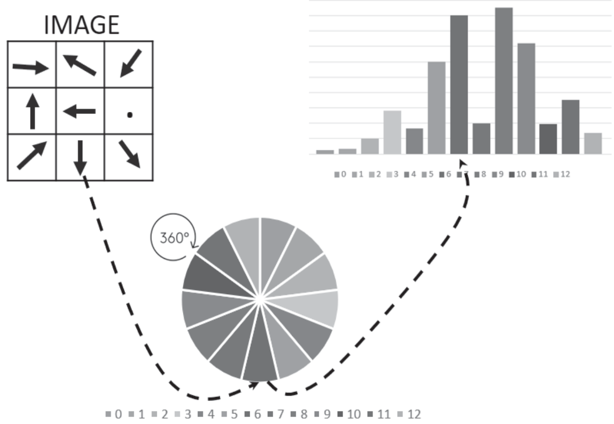

Fig. 3. Motion vector classification histogram (13 bins)

- α: a scale factor, it corresponds to the standard deviation of the Gaussian distribution.
- β: a shape parameter.

They find these parameters by using a maximum likelihood estimator of the GGD $(\widehat{\alpha}, \widehat{\beta})$ these estimators are defined by $x = (x_1, x_2, x_3, ..., x_L)$, and by the coefficients of residual data in one frame. Supposing that, each x_i is independent and L is the number of blocks per frame, and the digamma function is $\Psi(t) = \frac{\Gamma'(t)}{\Gamma(t)}$. [18] demonstrated that $(\widehat{\alpha}, \widehat{\beta})$ is the unique solution of the following equations:

$$\begin{cases} \hat{\alpha} = (\frac{\hat{\beta}}{L}\sum_{i=1}^{L}|x_i|)^{\frac{1}{\hat{\beta}}} \\ 1 + \frac{\Psi(\frac{1}{\hat{\beta}})}{\hat{\beta}} - \frac{\sum_{i=1}^{L} x_i^{\hat{\beta}} log|x_i|}{\sum_{i=1}^{L}|x_i|^{\hat{\beta}}} + \frac{log(\hat{\beta}\frac{\sum_{i=1}^{L}|x_i|^{\hat{\beta}}}{L})}{\hat{\beta}} = 0 \end{cases} \tag{5}$$

All data composing the signature (*Sign*) are extracted, where *i* represents the video's frame, the signature's vector represents a video sequence (*V*) is:

$$Sign_{V_i} = (Direction_i, Class_i, Intensity_i, \alpha_i, \beta_i) \tag{6}$$

C. Similarity measurement

The Principal Component Analysis (PCA) is used to reduce the dimension of data while maintaining up to 90% of its performance. Thereafter they chose Bounded Coordinate System as a model for measuring similarity, Zi Huang, Heng Tao Shen, Jie Shao, Xiaofang Zhou and Bin Cui [19] propose this model that captures the frame content distribution, which dominates the tendency of content.

Let *X* and *Y* videos, $BCS(X) = (O_X; \ddot{\Phi}_{X_1}; \ddot{\Phi}_{X_2}; ...; \ddot{\Phi}_{X_d})$ and $BCS(Y) = (O_Y; \ddot{\Phi}_{Y_1}; \ddot{\Phi}_{Y_2}; ...; \ddot{\Phi}_{Y_d})$, to calculate the similarity between BCS(X) and BCS(Y), two distances will be calculated. Where $d^X = d^Y$, $\|O_X - O_Y\|$ is the distance between two origins by translation, and it indicates the global difference between two sets of frames representing the video clips, and the average difference of all the content-changing indicated by the distance between each pair of bounded axes by rotation $\left\|\ddot{\Phi}_{X_i} - \ddot{\Phi}_{Y_i}\right\|/2$, but if $d^X > d^Y$, a scaling distance $\left\|\ddot{\Phi}_{X_i}\right\|/2$ will be added to a translation and rotation distance, therefore, the rotation and scaling indicate the content tendencies. The length of bounded principal component $\left\|\ddot{\Phi}_i\right\|$ is $2c\sigma_i$ [19].

$$D(BCS(X), BCS(Y)) = \lambda_0 \|O_X - O_Y\| + (\sum_{i=1}^{d^Y} \lambda_i \left\|\ddot{\Phi_{X_i}} - \ddot{\Phi_{Y_i}}\right\| + \sum_{d^Y+1}^{d^X} \lambda_i \left\|\ddot{\Phi_{X_i}}\right\|)/2 \tag{7}$$

They adapt these elements of equation (8) to converge on coherent results for study the trend of classification of classes for every amount of data.

$$D(BCS(X), BCS(Y)) = \lambda_0 \|O_X - O_Y\| + (\sum_{i=1}^{d^Y} \lambda_i \left\|\ddot{\Phi_{X_i}} - \ddot{\Phi_{Y_i}}\right\| + \sum_{d^Y+1}^{d^X} \lambda_i \left\|\ddot{\Phi_{X_i}}\right\|)/2 \tag{8}$$

Therefore, if $D(BCS(X), BCS(Y)) = 0 \Leftrightarrow BCS(X) = BCS(Y)$, that means *X* and *Y* are similar, and if the distance is bigger, *X* and *Y* are less similar. In this step, they add a λ coefficient based on genetic algorithm to avoid that certain values of the equation for being dominant. The genetic algorithm computation step takes a huge computing time and requires impressive system resources.

IV. Our approach

In BCMH framework there are two phases, the training step followed by a testing step. The training step includes hundreds of videos and consumes a high amount of time to analyse them, this is caused principally by the long processing time that consumes the genetic algorithms with its iterations in the first place, followed by PCA and signature extraction steps. Fig. 1 shows with (*) the phases of the BCMH framework that require a huge computation time. Upgrading the machine's performance will not solve the problem, and the time computation remains long.

In this scenario, the distributed video processing using Apache Hadoop is an interesting approach.

V. System Architecture and Implementation

To overcome the drawbacks of BCMH CBVR system, we propose to implement it on Apache Hadoop Framework. Compared to the local method in single machine [5] , the advantages using distributed processing based on Hadoop MapReduce are as follow: 1) We can divide the jobs processing of a huge video dataset into multiple sub tasks and assign them for different machines of the cluster. 2) The concept of running MapReduce jobs near the data, instead of moving the data to the processes can be also a key element. 3) The fault tolerance ensured by replicating data across multiple hosts allows to automatically control a single point of failure during program execution. 4) The ability to manipulate on various types of data, whether they are structured or unstructured like images, videos, large files, as is the case in the current approach. All these advantages and others favour the choice of Hadoop and make them the

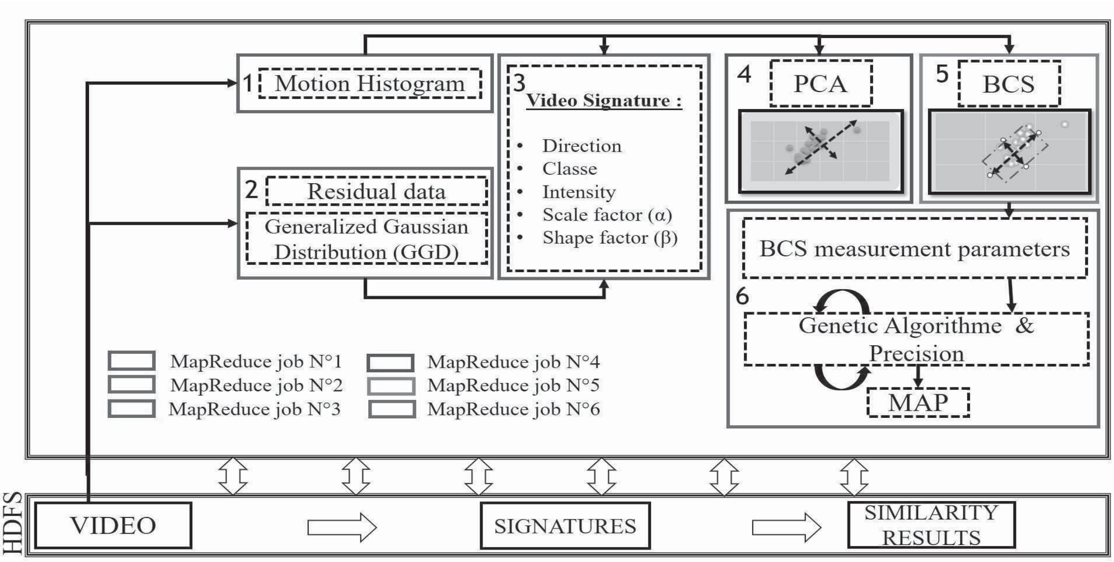

Fig. 4. BCMH process using Hadoop MapReduce distributed system

appropriate solution.

To deploy this CBVR framework on hadoop distributed system we wrote all BCMH programs with the MapReduce model according to the framework requirements.

In this paper, our solution is divided into two parts by using 6 chained Map-reduce jobs. First, signatures are extracted from each video stored on HDFS. and a signature database of the training videos is formed and stored on HDFS; secondly, we resize all videos signatures with PCA method, and to avoid that certain values of the signature being dominant, we randomly generate a series of individuals (λ) through an iterative process of genetic algorithm, on which we apply the BCS similarity measurement method, thereafter we keep the value of the individuals who have a better similarity result per class of video, in order to use them in the test phase.Fig. 4. presents the structure of the proposed scheme. The framework can analyze several videos at once, depending on the number of nodes available on the cluster.

A. Architecture of the distributed Hadoop system

In this paper, the experiment based on a typical Hadoop cluster as shown in Fig. 6, this cluster includes five nodes, one master and four slaves. These machines are connected to the same network segment. We started the solution with a Hadoop single node cluster, thereafter we increase number of nodes and compare the performance.

B. Bounded Coordinate of Motion Histogram freamwork: implementation on hadoop.

Fig. 5 describes the implementation steps of BCMH programs in Hadoop cluster using MapReduce.

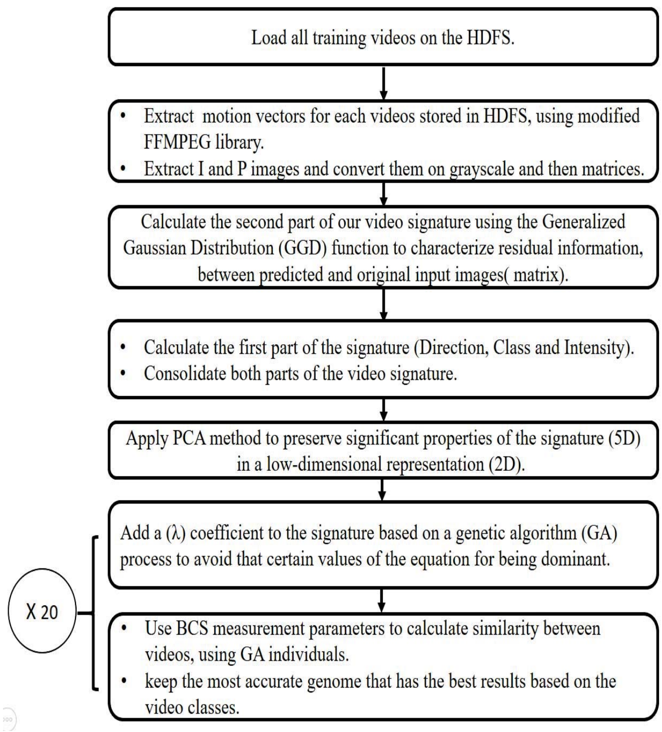

Fig. 5. Implementation steps of BCMH programs in Hadoop cluster

At the end of this step we have finished the training process and now have available the list of values on which the actual testing can be done.

VI. THE EXPERIMENTAL RESULTS AND ANALYSIS

In this section, we discuss and analyze the experimental results of the BCMH framework that we used with Hadoop

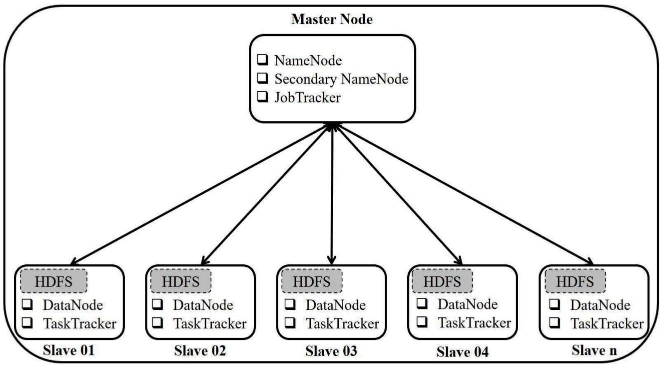

Fig. 6. Hadoop cluster

MapReduce programs on various cluster sizes, such as 1, 2 and 4 salve nodes.

A. Datasets

Fig. 7. HOLLYWOOD2 Human Actions and Scenes Dataset

HOLLYWOOD2 dataset contains 12 classes of human actions (AnswerPhone, DriveCar, Eat, FightPerson, GetOutCar, Handshake, HugPerson, Kiss, Run, SitDown, SitUp, StandUp) as shown in the Fig. 7, distributed over 1,707 video clips, more than 14.1 hours of video in total, and splited in two parts, training (823 videos) and testing (884 videos). The dataset intends to provide a comprehensive benchmark for human action recognition in realistic and challenging settings, extracted from 69 movies.

The idea behind choosing this dataset, that it presents different cases of the same action with different angles of capturing, let to generalize the results for each class.

B. Cluster setup

The experiments were evaluated on a cluster of 5 nodes (One Master and Four Slaves), on which we installed Hadoop 2.7.4 and FFMPEG 3.3 "Hilbert". Hardware specifications for the environment are described in Table I.

TABLE I. Hardware specifications

Name	Quantity	Description
Master	01	2 cores of Intel(R) Core(TM) T9900 @3.06GHz and 4GB RAM
Slaves	04	2 cores of Intel(R) Core(TM) T9900 @3.06GHz and 4GB RAM

C. Experimental Results

The following results represent a comparison between local and distributed system of the BCMH computation time in training step. Every "running time" is the average of five experiment attempts under the same circumstances. In Fig. 8, as follows we depict the comparison between the computation times, measured in minutes required to execute the three phases of our BCMH program previously described which demand a huge computation time (Signature extraction, PCA calculation and Genetic algorithm iterations), and it's clearly visible that the distributed system approach reduces the computation time on a 4 node Hadoop cluster to 76,75% of that of single machine.

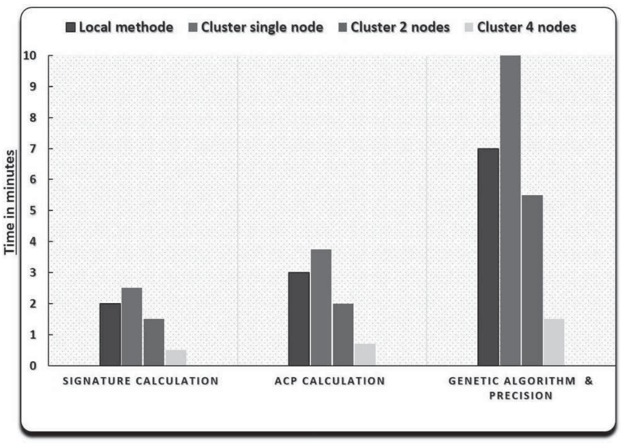

Fig. 8. Comparison of Execution Times

The second experiment takes the execution time as an observation variable and the number of input video files as well as the number of cluster slaves as control variables. Fig. 9. illustrates the experiment result. When the data set is very small, cluster with large number of nodes will not provide any improvement in computation time. Because, when data is very small Hadoop framework will take more time for division and synchronization of the tasks. We can remark from the graph that, as we increase the data set size,

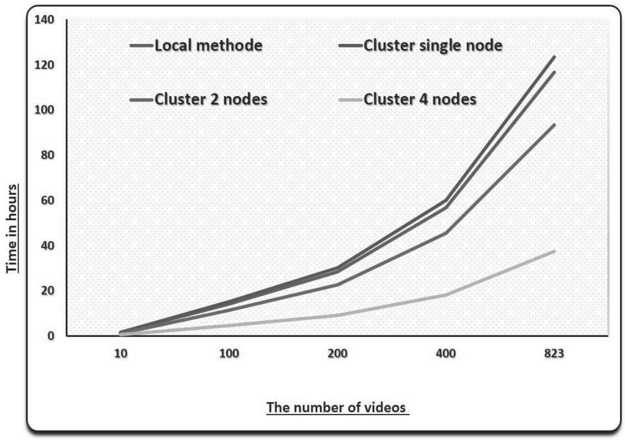

Fig. 9. Comparison of Execution Times

we get an important progress in the speed of execution of the BCMH algorithm. The total computation time on single machine grows linearly with increasing video files. The time of retrieval doesn't decrease linearly with the number of datanodes or data increase. That is caused by the time cost on initializing the MapReduce module, scheduling the task and clearing the output. Meanwhile, intermediate results generated by Map task also need some time to transfer to the Result task.

However, when we perform signature extraction and similarity measurement on a 4 node Hadoop cluster, the computation time is reduced to 32,25% of that of single machine. By adding more machines in our cluster, we can have more favourable results.

VII. Conclusion

This paper proposes an approach for distributed video processing on Apache Hadoop platform to fix the problem presented by the huge increase of the video data size. By utilizing Map-Reduce and HDFS across Hadoop clusters, the approach is able to handle large-scale of video data. To demonstrate the performance of the system, we chose training phase of a content based video retrieval system called BCMH as a use case. The experimental results shows that compared with the local method, using distributed system reduce the computation time. Analysis suggests that with more nodes added, the computation time can be also reduced.

During the testing phase of our use case, we noticed that a response to video similarity computation request takes a time interval that can be greatly reduced for better system reliability. Our future work will focus on how to reduce the computation time at this stage by using a free and open source distributed real-time computation system, that will contribute to making the BCMH solution more robust and scalable.

References

[1] Omnicore official website, YouTube by the Numbers, Web: https://www.omnicoreagency.com/youtube-statistics/.

[2] Statista official website, Social Media & User-Generated Content: Snapchat, Web: https://www.statista.com/topics/2882/snapchat/.

[3] QUELLEC, Gwénolé, LAMARD, Mathieu, CAZUGUEL, Guy, et al. "Real-time retrieval of similar videos with application to computer-aided retinal surgery.", *in Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE.* IEEE, 2011. p. 4465-4468.

[4] Snehal Anwekar, Isha Walimbe, Priyanka Suryagan, et al. "Flexible Content Based Video Surveillance System for crime Prevention based on moving object detection.", *in (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 7 (2)* , 2016, 655-660.

[5] EL OUADRHIRI, Abderrahmane Adoui, SAOUDI, El Mehdi, ANDALOUSSI, Said Jai, et al. "Content based video retrieval based on bounded coordinate of motion histogram.", *in Control, Decision and Information Technologies (CoDIT), 2017 4th International Conference.* IEEE, 2017. p. 0573-0578.

[6] T. White, *Hadoop: The Definitive Guide, M. Loukides, Ed.* O'Reilly, June 2009.

[7] C. Lam, *Hadoop in Action.* Manning Publications, December 2010.

[8] ANURADHA, J., et al. "A brief introduction on Big Data 5Vs characteristics and Hadoop technology.", *in Procedia computer science.* 2015, vol. 48, p. 319-324.

[9] BORTHAKUR, Dhruba, et al. *HDFS architecture guide.* Hadoop Apache Project, 2008, vol. 53.

[10] VAVILAPALLI, Vinod Kumar, MURTHY, Arun C., DOUGLAS, Chris, et al. "Apache hadoop yarn: Yet another resource negotiator.", *in Proceedings of the 4th annual Symposium on Cloud Computing.* ACM, 2013. p. 5.

[11] XIE, Jiong, YIN, Shu, RUAN, Xiaojun, et al. "Improving mapreduce performance through data placement in heterogeneous hadoop clusters.", *in Parallel & Distributed Processing, Workshops and Phd Forum (IPDPSW), 2010 IEEE International Symposium on.* IEEE, 2010. p. 1-9.

[12] SUNNY, Basil C., RAMESH, R., VARGHESE, Abraham, et al. "Map-Reduce based framework for instrument detection in large-scale surgical videos.", *in Control Communication & Computing India (ICCC), 2015 International Conference on.* IEEE, 2015. p. 606-611.

[13] LI, Moyun, YANG, Cheng, et TIAN, Jiayin. "Video selective encryption based on hadoop platform.", *in Computational Intelligence & Communication Technology (CICT), 2015 IEEE International Conference on.* IEEE, 2015. p. 208-212.

[14] TAN, Hanlin et CHEN, Lidong. "An approach for fast and parallel video processing on Apache Hadoop clusters.", *in Multimedia and Expo (ICME), 2014 IEEE International Conference on.* IEEE, 2014. p. 1-6.

[15] HUSAIN, Moula, MEENA, S. M., SABARAD, Akash K., et al. "Counting occurrences of textual words in lecture video frames using apache hadoop framework.", *in Advance Computing Conference (IACC), 2015 IEEE International.* IEEE, 2015. p. 1144-1147.

[16] SCHOEFFMANN, Klaus, LUX, Mathias, TASCHWER, Mario, et al. "Visualization of video motion in context of video browsing.", *in Multimedia and Expo, 2009. ICME 2009. IEEE International Conference on.* IEEE, 2009. p. 658-661.

[17] LAMARD, Mathieu, CAZUGUEL, Guy, QUELLEC, Gwénolé, et al. "Content based image retrieval based on wavelet transform coefficients distribution.", *in Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE.* IEEE, 2007. p. 4532-4535.

[18] VARANASI, Mahesh K. et AAZHANG, Behnaam. "Parametric generalized Gaussian density estimation." , *in The Journal of the Acoustical Society of America.* 1989, vol. 86, no 4, p. 1404-1415.

[19] HUANG, Zi, SHEN, Heng Tao, SHAO, Jie, et al. "Bounded coordinate system indexing for real-time video clip search." , *in ACM Transactions on Information Systems (TOIS).* 2009, vol. 27, no 3, p. 17.

[20] RATHORE, M. Mazhar, SON, Hojae, AHMAD, Awais, et al. "Real-time video processing for traffic control in smart city using Hadoop ecosystem with GPUs.", *in Soft Computing.* 2018, vol. 22, no 5, p. 1533-1544.

[21] JAI-ANDALOUSSI, Said, ELABDOULI, Abdeljalil, CHAFFAI, Abdelmajid, et al. "Medical content based image retrieval by using the Hadoop framework.", *In : Telecommunications (ICT), 2013 20th International Conference on. IEEE*, 2013. p. 1-5.