VOL. 10, NO. 13, JULY 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences


www.arpnjournals.com

VIDEO ANALYTICS USING HDVFS IN CLOUD ENVIRONMENT

Dhinakaran K., Silviya Nancy J. and Duraimurugan N.
Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha University, Chennai, India
E-Mail: maildhina.k@gmail.com

ABSTRACT

The proposed research work focuses on video analytics which is been emerged as the fastest growing research area in the schematic sketch of Big data and its applications in cloud. The exploration of video is been well marginalized and organized by Hadoop Infrastructure, whose propagation is discussed. Large volume of unstructured data is been produced in day to-day activities of the people, not exempting various organizations that are recorded and maintained eventually for research and other purposes. Traditional database architectures were not able to handle the generation, storage and access of this huge amount of unstructured (video) data. This proliferation led to the collaboration of video-analytics with a conservative database theory called Big Data. There are several structured database that are developed to master the needs for accessing and maintaining the videos. Despite, these developments, the retrieval and processing of the particular environment detail, specifically a featured human, or an object from the videos takes huge amount of time which is not too effective. So, this shortcoming is been forwarded to the Hadoop Distributed Video File System (HDVFS) whose Map-Reduce Framework process the recognition of stipulated image/object and their behavior from the unstructured video in the cloud repository which stores the intensive-sized files like Amazon S3 storage bucket.

Keywords: video analytics, big data, cloud storage, hadoop distributed file system (HDFS), amazon S3.

1. INTRODUCTION

The most leveraging and vibrant task that, the whole world of this era is engrossed in is, the ‘creation of data’. So, it is obvious that, there is exponential outburst in the growth of data which booms up swiftly in an unimaginable rate, overwhelming the Moore’s concept of transistors. Big Data – originates a dominating revolution in the IT industries, academics, and various other organizations. The assured benefits and even pitfalls also have started gaining attention globally. The phrase BIG DATA has become a buzzing word which facilitates the beneficiaries with collection of vast amount of data, its storage and giving a new form to scrutinize its appearance have reshaped the priority of using this technology in different fields. Manipulation of massive amount of data is a tedious and considered to be a complex task for the human beings to inculcate the specific information effectively.

The daily creation of data is probably about 2.5quintillionproduced by all the origins such as pictures, videos, mobile devices, social media sites, financial, manufacturing, marketing, physical and biological sciences and many more is what we usually define it as big data. It has both structured and unstructured volume of data. The processing of this vast data by traditional databases normally takes a longer time than usual and it automatically gets slower and it cannot be supported by any other software techniques also. The datasets of big data includes, multi-dimensional data, multi-modal data which is very huge and cannot be processed easily.

The concept of big data is endowed with certain characteristically defined attributes such as velocity, volume, variety, veracity. As the data management has become very crucial these days, the software industries have entertained several programming suites and models with data tools. These make the data processing little effective but these solutions make only a part. The big data has not only boomed in the field of industrial research but also have stamped its footsteps in astronomy and life sciences too. Whereas astrologically the job of astronomer is to capture the pictures relation to astronomy and in medicinal facts, it is to maintain the history of the patient databases.

The influence of big data has not left the education also. Here, in the view of educational reimbursement, with the compilation of all the statistics of various academic institutions, a better future education system can be built based on the information gathered on variety of chambers and its current developmental aspects. Big data offers its stability to various other causes including detailed analysis on transportation for efficient road network with visualization, monitoring and developing the environment, Computational sciences that help to dramatically scrutinize the population network, several financial systems and security management. Figure-1 depicts the work flow of the big data, on how the data is acquired for extraction to hold and carry out the analysis on variety of data.

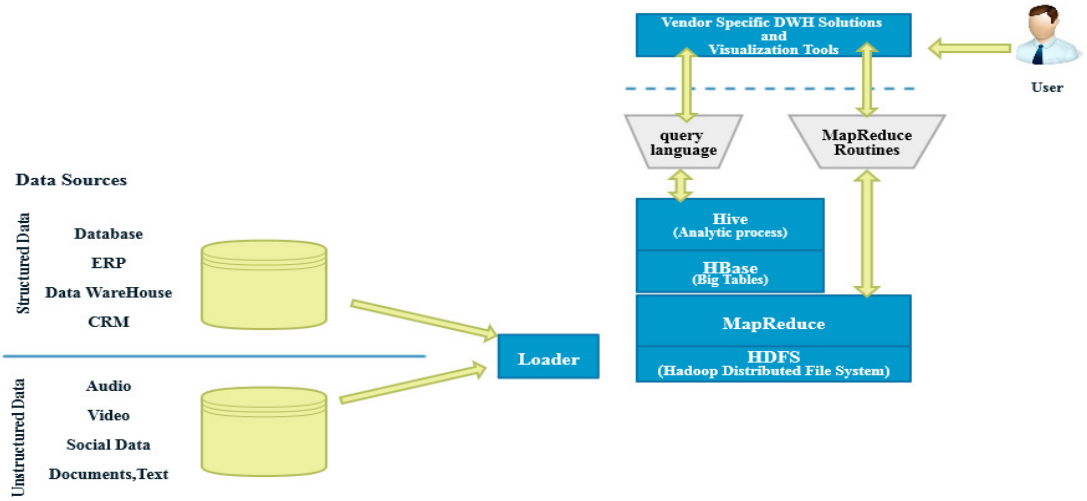

Figure-1. Pipeline processing of Big Data analytics with Hadoop MapReduce framework.

With improvements in quality and ambiguity of data, the term "big data" has reached its hype in effectively handling any type of structured, unstructured and semi-structured data. In today's scenario 90% of world's data is been created in recent years, out of which 80% is the constituent of unstructured video uploads and images. Not only the information that is in the form of text is accumulating rapidly, but the images and videos are also uploaded and used massively by the users in regular basis which occupies excessive area of storage capacity. Even though, the big data nurtures only the well-organized data like Facebook, twitter, etc.., its necessity is also been demanded in the market of evolving video analytics.

The digital devices flash millions of videos worldwide for every second. Moreover, the online archives are flushed with the new footages that are uploaded by the users for every minute. This does not end here, whereas there are other sources from surveillance cameras exceeding in their count and memory sizes. With this persistent growth, the analysis of videos takes up a new identity in the engineering and technological advancements. To define it approximately, video analytics is culmination or pulling of relevant images and information from the cluster of digitalized videos. It is most commonly used for regulation administration and intelligence.

The research on video analytics also needs the optimality of big data. The amount of pictures (images) and videos that are being uploaded are increasing widely. Nearly 5 billion videos are being uploaded in the Facebook per month. But managing and manipulating of these video-oriented data has become a tedious task day by day since the amount is increasing proportionally. Video-analytics offers many privileges in handling sophisticated images in pixel by pixel basis, to acquire the innermost details from the video through the processing image. The procedure of video analytics is shown in Figure-2.

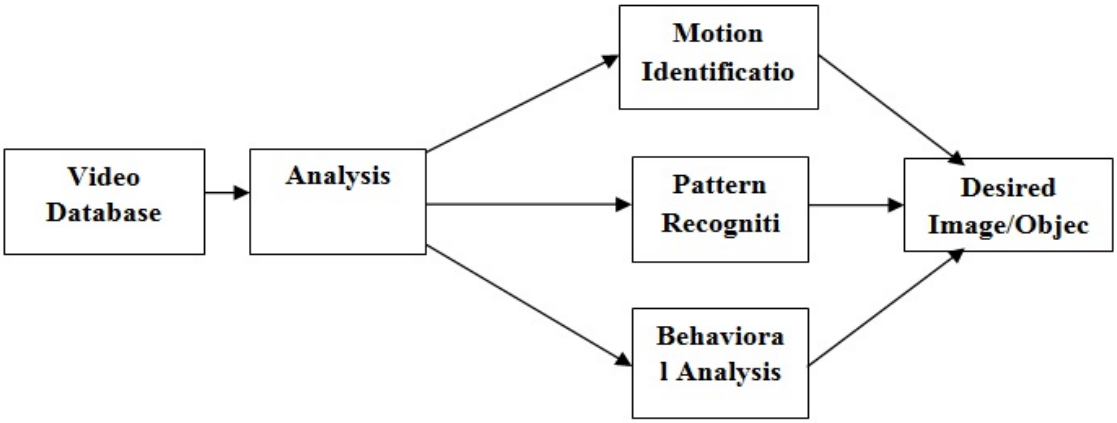

Figure-2. Sequence of steps in video-analytics.

There are many images and video processing as well as computer vision algorithms to operate on, but the computational power of those algorithms are limited to certain extent. The frameworks and algorithm in machine learning intelligence, pattern recognition and computer vision enables to perceive the object and behavior. But, the significance and importance of video analysis grows eventually, which needs an even better tool to master the needs and to provide progression in horizon for safety and security. Usually, the processing is done by parallel assignment of tasks to the computers connected in the network. These require intensive resources for performing a particular task like processing of parameters like color, pixels, etc. To customize all these tasks a frame work called Map-Reduce is been introduced which stabilize the parallelizing of scalable and efficient simulation of resource-intensive images and videos.

A report on current statistics of data creation reveals that 90% of world's data is been generated in last two years. A survey by IBM proclaims that, 80% of data produced data is unstructured with data gathered from social networking sites, sensor gathering climatic conditions and control, pictures, videos, are some of the examples. This paved way for the advent of "Big Data Theory". To manipulate this huge amount of data which is not formatted became a tedious task with the limited storage-capacity databases. Earlier, in 90's search engines and indexes were maintained for operating on the information retrieval, and another case it was returned by humans. In recent years, all the organizations started booming up with data creation especially with the web pages to hold tons and tons of data. This originated to reach the results sooner and, this was also relied on web-based search engines where the processing would be carried out in a distributed way, so that results would be returned faster.

With the drastic increase in volume and variety of data there was a need for optimal solution, which gave rise to a new technology, called "Hadoop". The most acceptable advantage is that, this framework can handle large amount of data very faster and yield the results quickly which was developed to overcome the shortcomings of relational databases. Currently, the organizations and many IT infrastructures are in need of a methodology to process these unstructured heavy datasets. So, Hadoop remains as an efficient solution for this crisis and provides proficiency in analytics. Instead of storing the big data in the central database, it would maintain n number of clusters. These clusters can be associated with the multiple machines (each of them with 2 – 8 CPUs). The Hadoop Cluster structure is mainly designed for managing the huge datasets, which is distributed among different machines to complete the task earlier. A single Hadoop instance consists of Name node along with number of Data nodes.

The two major components of Hadoop are Hadoop Distributed File System (HDFS) and Map Reduce. The HDFS is a well-organized file system and it is java-based which can store and handle data without prior organization. Map Reduce is called as "Job Tracker", which associates Hadoop by scaling, distributing and parallelization of services. It has two phases called "Map" and "Reduce" phase. During Map phase the given input from the Hadoop Distributed File System (HDFS) will be split into chunks, which would be made to run parallel in the Hadoop cluster. In turn, during the reduce phase, consolidated outcome is arrived and stored in HDFS.

With the exploration mentioned above, the proposed system is based on the concept of video analytics in the Hadoop Framework for detecting and analyzing particular object from the video. From the cloud database, the videos can be retrieved, where these videos would be split into n number of frames based on the necessity. These divided frames will be fed into Hadoop Distributed File System (HDVFS), which will serve as input to the mapping phase. And in reducing phase the object will be identified based on the HDVFS proposed algorithm.

The research work consists of related works of literature in Chapter – II, Chapter – III holds the explanation of proposed HDVFS algorithm and model and Chapter IV is engraved with experimental results and Chapter – V is furnished with conclusion and future work and noted with References.

2. RELATED WORK - BACKGROUND

This chapter is completely based on the literature review of big data and Hadoop Framework. In this subsequent discussion various challenges in big data are enlightened. Adding to that, the integration of big data and Hadoop MapReduce Framework is also been compared and evaluated. More significantly the deliverable featured functionality and conceptual algorithms for different tracking and recognition of objects from the cluster of images or videos using video analytics is included in the brief elaboration.

Myoungjin Kim *et al.* 2013, proposes the Hadoop Distributed Video Transcoding System in the Cloud Computing environment. In this system, transcoding of the one video codec format into diverse formats for various devices like smartphones, computers, etc. This is been implemented on Hadoop MapReduce Framework for fast processing of transcode of video codices' effectively. Performance evaluation was carried out on 28-node cluster which promises tremendous speed and quality. The authors S. Zhang *et al.* 2012, discusses on the architecture for video surveillance system in assistance with video analytics.

The features of video analytics are used for obtaining high quality videos and perform real-time transcoding, compression and security with video-analytic algorithm. Video analytics includes intelligent tasks like motion detection, human identification, pattern recognition, etc. There are various methodologies that were proposed earlier like point tracking, kernel tracking and many others, but the authors Tao LUO *et al.* 2014, focuses and discusses on silhouette tracker which identifies the objects easily. The presenters describe novel block based segmentation with silhouette tracker for

finding the object. The owners of this article Kalpana Dwivedi and Sanjay Kumar Dubey (2014) express various view and analytical review on Hadoop Distributed File System. It obviously provides detailed explanation on the components of Hadoop like HDFS and MapReduce. And also in addition to that various Hadoop oriented tools like HIVE, Hbase, Chukwa, etc., and their file processing segments are also conferred.

Amrit Pal and Sanjay Agrawal (2014), talk about the big data challenges that were extended to the data analysts. This research work defines the importance of Hadoop Framework for their MapReduce jobs. In this paper, they have provided a brief explanation on memory limits that are used for mapping and reducing tasks. Walisa Romsaiyud and Wichian Premchaiswadi (2013), the motivation of this article is placed on the interaction of MapReduce tasks with machine learning algorithms. Also overcomes the time-consuming problem of artificial learning based algorithms and optimizes the tasks by using extended MapReduce execution methods which minimizes the cost, and it was experimented in EC2 Hadoop with 20-nodes cluster. Shweta Pandey and Vrinda Tokekar (2014), discusses on the importance of big data in various domains and the prominent revolution it has made the data collection jobs. Furthermore, the emphasis on MapReduce for handing the big data is also been clearly elaborated in the survey.

The several scheduling algorithms and brief tabulated explanation on how MapReduce jobs can be done effectively is also been added content to that. Hadoop MapReduce has gained a lot of attention nowadays due to the exploring big data progression. A precise example of the above mentioned framework is word count which is explained in the article Amrit Pal and Pinki Agrawal (2014), where the authors contribute an analysis on large data sets experimented on Hadoop with the in-built default Hadoop Distributed File System (HDFS). The investigational model includes mapping and reduce jobs based on the bytes of the file and increases the performance. Shen Li *et al.* 2014, proposes WOHA workflow model which delivers an effiecient reducing jobs. This model allows the client to summarize own scheduling plans which is later used for resource handling by the master node. The jobs are assigned based on the priorities and the resulting performance is improved 10% comparing to the existing methodologies. E. Dede *et al.* 2014, presents a brief conversation on usage of NoSQL with the Hadoop MapReduce framework.

This paper introduces the pipeline streaming of non-executables in MapReduce framework with Cassandra data sets and reports the performance with other streaming techniques. Also includes the scheduling phase for loading the datasets for execution in appropriate time rather using local database servers. Due to the increase in data to unpredictable rates, the authors Mehmet C.Okur and Mustafa Buyukkececi (2014), insists on the inception of these revealing concepts in educational organizations too, which would lead to the development of engineering sciences.

The literature survey conveys the significance of the Hadoop MapReduce Framework in efficiently handling the large data sets. And it has been derived that it will also relieve the heavy processing of video manually or with less-efficient tools. Instead, Video-Analytics can be incorporated with the Hadoop Architecture including HDFS and MapReduce which would provide faster results, and it has been discussed in the proposed system.

3. RECOMMENDED PROPOSED SYSTEM FOR HDVFS

The hype created by big data has reached all the organization for encountering the huge data sets which was mostly text-based sets. The video seems to be the next emerging source for various industrial and academic concerns for knowing better perception of the consumers. With the electronic devices like digital cameras, surveillance footages, mobiles and many others, the eruption of unstructured data is evolving rapidly plugging the video databases and online archives.

The videos may be of different types such as Social Network, Media Sharing and Organization surveillance videos. Several methodologies in object detection, pattern recognition, and computer vision contributed a lot to enlighten the video streaming and motion detection to perfect level of security. In the proposed system, the difficulty in manually processing of videos is mastered by Hadoop Distributed Video File System (HDVFS) with which the desired part or object is detected by MapReduce framework. Here the videos are collected and maintained in the video database (Amazon Storage S3), which is manipulated through the HDVFS which consists of various data notes for handling the task of identification of particular object in the image which is done by Mapping and Reducing. Then the user can avail the required object or motion from the images which will be specified with frame number and the video in which it was present. The architecture is elaborated in Figure-5; the workflow of the system is defined in next chapter.

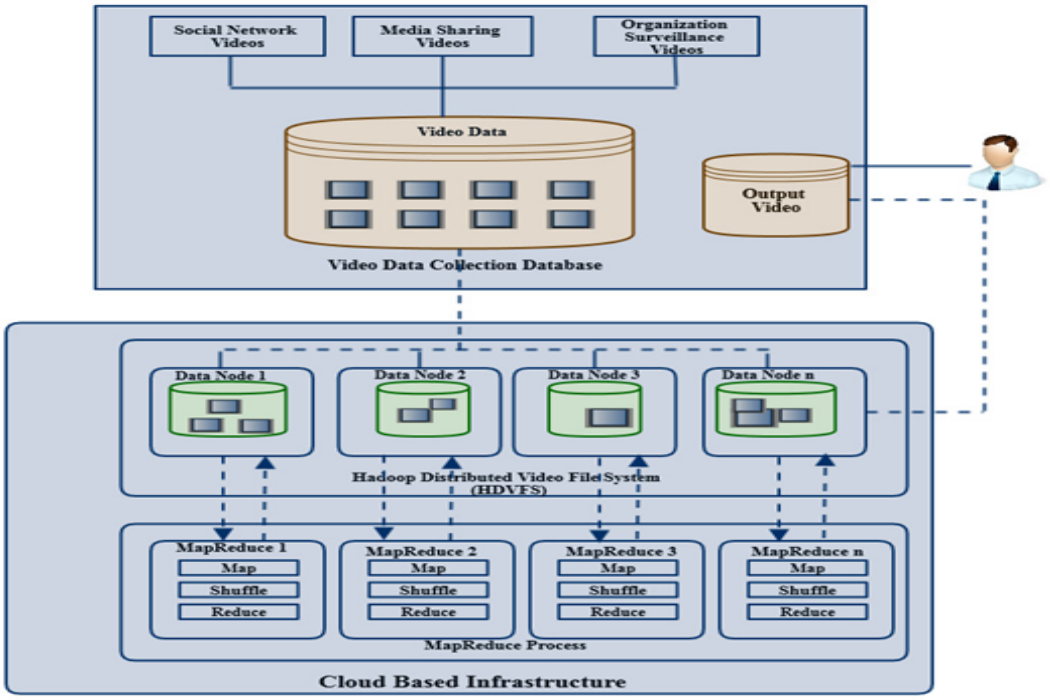

Figure-3. Proposed model for Hadoop distributed video file system (HDVFS).

a) Workflow description for Hadoop distributed video file system (HDVFS)

The working of the HDVFS is elucidated in the trailing Figure-4 below.

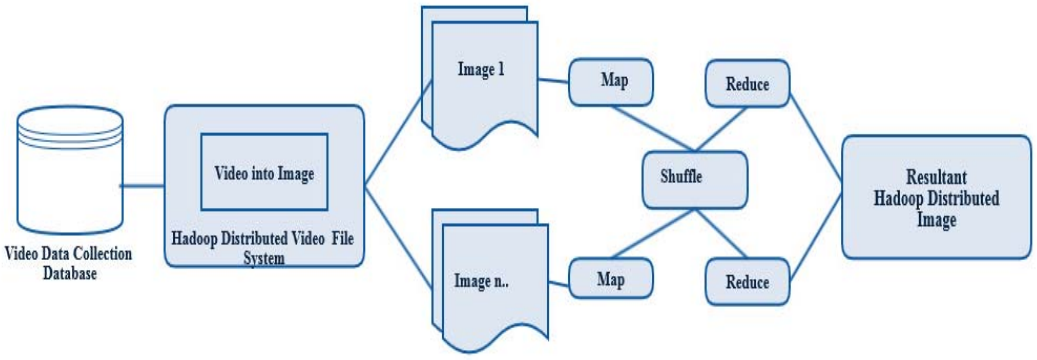

Figure-4. Picture of MapReduce framework using HDVFS.

1) Analyze the frame that has to be given as a result.
2) Identifies the collection of videos in the database.
3) Passed in HDVFS, where all the frames are mapped.
4) If the required frame is found, then it will be stored in the database for further use.

The above mentioned steps are explicitly depicted above, which is also accompanied by the algorithmic flow of statements.

HDVFS algorithm

Video database (v)
User input(i)
Step-1: Checks the number of videos from 1 to n.
Step-2: In HDVFS , Videos are converted into frames
Step-3: all frames are forwarded to MapReduce process

MapReduce algorithm for video frame

Function map (float frame, double videoDB)

```
        //frame: Video frames
        //videoDB: Video Database
For each frame f in videoDB ;
Emit(f,1)
Function Reduce(float frame, Iteration Count)
        //frame: a frame
        //counts: a list of arranged frame counts
Sum=0
For each c in Count;
Sum+ = ParseInt(c)
Emit(frame, sum)
```

Step-4: find the related output image then send to VideoDB.

Step 5- Output (Video number with frame number)

4. RESULTS AND DISCUSSIONS

The Video Analytics is usually processed through series of steps;

Online experimentation setup

a) Storage in cloud

The experimentation setup for video storage is been furnished with Amazon S3, which is been carried out by creating the instance along with the initialization of bucket storage where files (videos) of any size can be uploaded and saved.

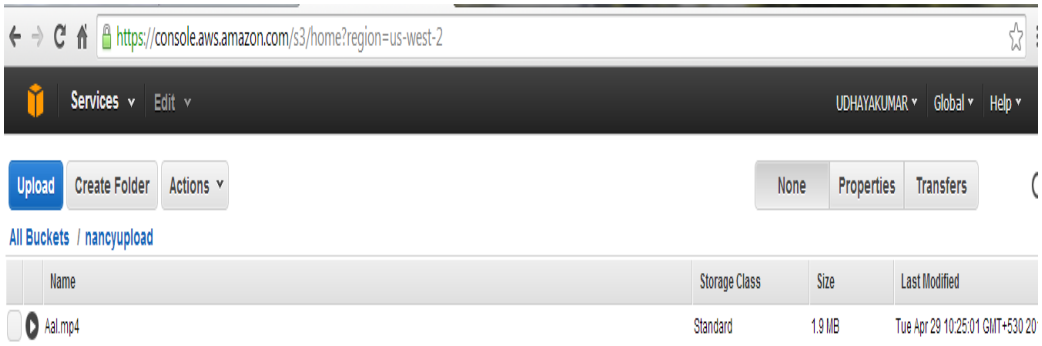

Figure-5. Bucket storage in Amazon S3.

Offline experimentation setup

a) Videos to frames / images

The videos are retrieved from the cloud storage, where the video is been made to split into *n* number of frames/images which is then framed as input to the Hadoop Distributed Video File System (HDVFS). The sample analysis of video to frame conversion is been carried out by the tool whose specifications and results are shown below.

Table-1. Specifications of video to frames/images conversion.

S. no.	File name (Video File)	File size (MB)	File type (Format)	Video to frame format	File (Image) size (MB)
1	video	25	.wmv	.bmp	2.63
2	good	30	.mp4	.bmp	2.92
3	scenery	35	.avi	.bmp	4.25
4	myvideo	40	.wmv	.bmp	4.85
5	baby	20	.avi	.bmp	2.10

The Table-1 specifies the sample instances of videos whose frames are been split and their respective file sizes are displayed above. The Offline setup shown in Figure-6 & Figure-7, exhibits the dispersal of split image. As a brief note, Figure-6 presents how the frames are captured while video is been streamed. Figure-7 represents the saved frame (bmp).

Figure-6. Frame captured during streaming.

Figure-7. Saved frame.

b) Mapping phase

Next is the mapping phase, the images in the HDVFS are segregated with the assistance of the data nodes in the file system using the proposed HDVFS algorithm. At this stage, numerous images will be compared

c) Reducing phase

With effective comparison from by the mapping phase, the required image or an object can be detected in reduce phase based on the two key concepts like Motion Identification, which is determined for investigating each and every pixel and even their slightest movements and change in posture of object. In addition to that, recognition the similarity in patterns is differentiated and analyzed in all the frames as shown in Figure-8.

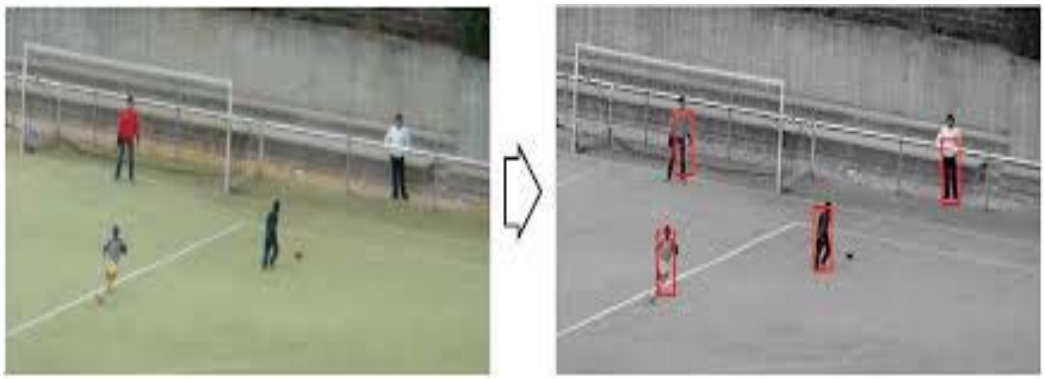

Figure-8. Detection of images after mapping and reducing.

5. CONCLUSION AND FUTURE WORK

Video analytics is considered as the emerging phenomenon in today's world. It offers various advanced and useful services and minimizes the manual work and reduces the time taken for analyzing and tracking the objects. In the proposed work, we have recommended a Hadoop Distributed Video File System (HDVFS), where Object Tracking is made simple and efficient with MapReduce framework. The experimental study also prescribes the online storage of videos shown in Amazon S3 bucket storage. Further, as a future implementation the Hadoop Video MapReduce can also be implemented in public or private cloud.

REFERENCES

[1] Myoungjin Kim, Yun Cui, Seungho Han and Hanku Lee, "Towards Efficient Design and Implementation of a Hadoop-based Distributed Video Transcoding System in Cloud Computing Environment", International Journal of Multimedia and Ubiquitous Engineering Vol. 8, No. 2, March, 2013.

[2] S. Zhang, S. C. Chan, R. D. Qiu, K. T. Ng, Y. S. HungAnd W. Lu, "On the Design and Implementation of a High Definition Multiview Intelligent Video Surveillance System",IEEEInternational Conference on Signal Processing, Communication and Computing (ICSPCC), 2012.

[3] Tao LUO, Ronald H. Y. Chung, K. P. Chow, " A Novel Object Segmentaion Method for Silhouette Tracker in Video Surveillance Application", International Conference on Computational Science and Computational Intelligence, 2014.

[4] Kalpana Dwivedi and Sanjay Kumar Dubey, "Analytical Review on Hadoop Distributed File

System", IEEE, Confluence the Next Generation Information Technology Summit (Confluence), 2014.

[5] Amrit Pal and Sanjay Agrawal, "An Experimental Approach towards Big Data for Analyzing Memory Utilization on a Hadoop cluster using HDFS and MapReduce", IEEE International conference on Networks and Soft Computing, 2014.

[6] Walisa Romsaiyud and Wichian Premchaiswadi, "An Adaptive Machine Learning on Map-Reduce Framework for Improving Performance of Large-Scale Data Analysis on EC2", Eleventh International Conference on ICT and Knowledge Engineering, 2013.

[7] Shweta Pandey and Vrinda Tokekar, "Prominence of MapReduce in BIG DATA Processing", Fourth International Conference on Communication Systems and Network Technologies, 2014.

[8] Amrit Pal and Pinki Agrawal," A Performance Analysis of MapReduce Task with Large Number of Files Dataset in Big Data Using Hadoop", Fourth International Conference on Communication Systems and Network Technologies, 2014.

[9] Shen Li, Shaohan Hu, Shiguang Wang, LuSu, Tarek Abdelzaher, Indranil Gupta, Richard Pace, "WOHA: Deadline-Aware Map-Reduce Workflow Scheduling Framework over Hadoop Clusters", IEEE 34th International Conference on Distributed Computing Systems, 2014.

[10] E. Dede, B. Sendir, P. Kuzlu, J. Weachock, M. Govindaraju, L. Ramakrishnan, "A Processing Pipeline for Cassandra Datasets Based on Hadoop Streaming", IEEE International Congress on Big Data, 2014.

[11] Mehmet C.Okur and Mustafa Buyukkececi, "Big Data Challenges in Information Engineering Curriculum", IEEE, 2014.

[12] Ge Song, ZideMeng, FabriceHuet, FredericMagoules, Lei Yu, "A Hadoop MapReduce Performance Prediction Method", IEEE International Conference on High Performance Computing and Communications & IEEE International Conference on Embedded and Ubiquitous Computing, 2013.

[13] Megha Sharma, NitashaHasteer, AnupriyaTuli, AbhayBansal, "Investigating the Inclinations of Research and Practicesin Hadoop: A Systematic Review", IEEE,2014