Content Based Video Retrieval by Using Distributed Real-Time System Based on Storm

Abderrahim Sekkaki, University Hassan II of Casablanca, Casablanca, Morocco
Said Jai Andaloussi, University Hassan II of Casablanca, Casablanca, Morocco
Abderrahmane Adoui El Ouadrhiri, Hassan II University of Casablanca, Casablanca, Morocco
Othmane El Warrak, University Hassan II of Casablanca, Casablanca, Morocco

ABSTRACT

Time processing is a challenging issue for content-based video retrieval systems, especially when the process of indexing, classifying and retrieving desired and relevant videos is from a huge database. A CBVR system called bounded coordinate of motion histogram (BCMH) has been implemented as a case study. The BCMH offline step requires a long time to complete the learning phase, and the online step falls short in addressing the real-time video processing. To overcome these drawbacks, this article presents a batch-oriented computing based on Apache Hadoop to improve the time processing for the offline step, and a real-time oriented computing based on Apache Storm topologies to achieve a real-time response for the online step. The proposed approach is tested on the HOLLYWOOD2 dataset and the obtained results demonstrate reliability and efficiency of the proposed method.

KEYWORDS

CBVR, Distributed Processing, Hadoop, Real Time, Storm, Stream Processing, Time Processing, Video Processing

INTRODUCTION

Over the last decade, video data has increased dramatically, especially with the development of smartphones and the evolution of social media. As an example, the statistics published by Facebook, about the 4th quarter of 2018 indicates that, there are over 2.32 billion monthly active users (Facebook, 2018), watching 100 million hours of video every day (Smith, 2019). Sometimes, the same video content is posted or shared several times, and since the storage range is limited, it is worthwhile to search for similar content to delete it, use a quota, or classify it. Since the large volume of data leads to a difficult retrieve, the approach and the infrastructure used will certainly help to curb some difficulties. Various Content-Based Video Retrieval systems (CBVR) that allow efficiently retrieving similar videos to a query video from a database have been developed. Quellec et al. (2011) proposed

DOI: 10.4018/IJERTCS.2019100104

Copyright © 2019, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
a CBVR system for real-time retrieval of similar videos with application to computer-aided retinal surgery. They use Dynamic Time Warping (DTW) technique to measure distance between video subsequences. Loukas (2018) presented a content-based video analysis of surgical operations. He reviews recent developments and analyzes future directions in the field of content-based video analysis especially on surgical operations. Ishtiaq et al. (2018) proposed a method that receives video content and metadata associated with video content. The method then extracts visual, audio, and textual features of the video content based on the metadata. A set of video segments of the video content is identified based on the composite features of the video content. Thereafter the segments will be identified based on a user query. A CBVR system called Bounded Coordinate of Motion Histogram (BCMH), with offline and online processing parts is presented by El Ouadrhiri et al. (2017). To characterize videos, they used vector motions and residual data, and to calculate the similarity, they used the Bounded Coordinate System (BCS). The similarity measurement accuracy of BCMH system is interesting, but the offline processing step requires a long time to complete the learning phase with large videos. Likewise, the online processing part is unable for similarity measurement in real-time. Therefore, optimizing the BCMH computation time is the challenge in the current scenario. In this paper, the authors propose a batch-oriented computing based on Apache Hadoop to improve the time processing during the offline step of the BCMH system. The batch processing is, in general, efficient in processing large volumes of data, and Apache Hadoop is the most common framework used. In addition, a real-time oriented computing based on Apache Storm topologies was proposed to achieve the real-time response for the online step.

The outline of this paper is as follows. In the section II, authors introduce the basic concepts and related work of CBVR systems using Hadoop distributed platform and the real-time video processing using Apache Storm. Section III presents materials and methods. The next section defines the approach. Section V describes the architecture and the implementation of the system. Section VI presents the experimental results and analysis. Finally, Section VII concludes this paper.

BACKGROUND

Apache Hadoop is an open-source framework invented by Doug Cutting and Mike Cafarella in 2005, it allows the distributed processing of large data sets through clusters of computers using simple programming models (Anuradha, 2015). It offers reliable, scalable and distributed computing. The main technology of Hadoop is HDFS, YARN and Map Reduce. Hadoop Distributed File System (HDFS), is a distributed file system for big data and large files, that provides high-throughput access to application data, it achieves reliability by replicating the data across multiple hosts (Borthakur et al., 2008). 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 (Vavilapalli et al., 2013). Hadoop Map Reduce is the programming model of Hadoop, which includes Map function and Reduce function for parallel processing of large data sets (Xie et al., 2010). Figure 1 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).

In this paragraph, authors present some of the relevant and recent studies discussing approaches for video processing on Apache Hadoop. Kesavaraja and Shenbagavalli (2018) proposed an intelligent video splitter which uses the map reduce algorithms to provide efficiency based on time factor. This process confirms that splitting the video into nearer 64MB blocks with an increase in the number of mappers, reduce the processing time. Parsola et al. (2017) proposed an approach for storage and processing of video data using Hadoop/MapReduce. They detect moving objects from videos. Frames are extracted from video for motion detection, then converted into sequence file. In sequence file, frames are merged in a key value pair where key is the name of the frame and value is the contents of the frame. Thereafter, sequence or input file is split into blocks of 64 MB before it can be processed.
Integrating Caching Techniques in CDNs using a Classification Approach
www.igi-global.com/article/integrating-caching-techniques-cdns-using/1463?camid=4v1a

Replication Strategies for Video On-Demand over Wireless Mesh Networks: A Cross-Layer Optimization Approach
www.igi-global.com/chapter/replication-strategies-video-demand-over/65677?camid=4v1a