

I. INTRODUCTION

The growth of the amount of image data produced on a daily basis in social media forces the adaptation of traditional image analysis and indexing approaches towards scalable solutions. Image searching is one of the most important services that need to be supported by such systems. One way to retrieve target image from image collections is called Content-Based Image Retrieval (CBIR) systems. In the CBIR systems, image processing algorithms are used to extract feature vectors that represent image properties such as color, texture, and shape. It is possible to retrieve images similar to one chosen by the user (query-by-example). The main advantages of this approach is the possibility of an automatic retrieval process, contrasting to the effort needed to annotate images. Although the result in terms of accuracy is well done but we found that a computation time of the two main processes of CBIR System is time-consuming, laborious and expensive. Nowadays, numerous researchers attempt to solve this problem by applying the distributed computing methods. The approach includes different variety of areas such as, cluster computing in order to decrease the computational time [1] [2] [3]. Kao, et al [2] offered a cluster platform which was compatible with running other retrieval approaches practiced in CBIR systems. Their study presented the rudimentary standards of image retrieval with dynamic feature extraction by means of cluster platform architecture. Yongquan Lu, et al [1] suggested a parallel technique to execute similarity comparison and feature extraction of visual features rooted in cluster architecture. The subsequent conducted experiments indicated that a parallel computing technique can be used with the intention of considerably improving the performance of a retrieval system. The major emphasis in [2] is workload balancing through the cluster with execution performance and scheduling heuristic measurements of the launched prototype. Consequently, cluster computing lessens the computational time overall. However, in order to define a distributed algorithm and program it with cross-platform ability is a crucial mission [4]. In this study, we propose a parallel processing technique via Hadoop MapReduce framework [5] [6] [7] for content-based image retrieval system instances. Accordingly, the two contributions discussed in this study are Index creation process and similarity measurement process through MapReduce.

This paper is organized as follows. Section II presents a background of CBIR system. Sections III address the basic concept of Hadoop MapReduce. Section IV presents an applying MapReduce scheme to image processing. Section V proposed the Hadoop MapReduce Framework for Content-Based Image Retrieval Systems. Section V provides a summary and conclusion