Information Sciences 263 (2014) 60–86

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Efficient and robust large medical image retrieval in mobile cloud computing environment

Yi Zhuang [a,*], Nan Jiang [b], Zhiang Wu [c], Qing Li [d], Dickson K.W. Chiu [e], Hua Hu [f]

[a] College of Computer & Information Engineering, Zhejiang Gongshang University, Hangzhou, PR China
[b] Hangzhou First People's Hospital, Hangzhou, PR China
[c] Jiangsu Provincial Key Laboratory of E-Business, Nanjing University of Finance and Economics, Nanjing, PR China
[d] Department of Computer Science, City University of Hong Kong, Hong Kong Special Administrative Region
[e] Faculty of Education, The University of Hong Kong, Hong Kong Special Administrative Region
[f] School of Computer, Hangzhou Dianzi University, Hangzhou, PR China

ARTICLE INFO

Article history:
Received 5 April 2013
Received in revised form 16 September 2013
Accepted 13 October 2013
Available online 21 October 2013

Keywords:
Medical image
Multi-resolution
Mobile cloud

ABSTRACT

This paper presents an efficient and robust content-based large medical image retrieval method in mobile Cloud computing environment, called the MIRC. The whole query process of the MIRC is composed of three steps. First, when a clinical user submits a query image I_q, a parallel image set reduction process is conducted at a master node. Then the candidate images are transferred to the slave nodes for a refinement process to obtain the answer set. The answer set is finally transferred to the query node. The proposed method including an priority-based robust image block transmission scheme is specifically designed for solving the instability and the heterogeneity of the mobile cloud environment, and an index-support image set reduction algorithm is introduced for reducing the data transfer cost involved. We also propose a content-aware and bandwidth-conscious multi-resolution-based image data replica selection method and a correlated data caching algorithm to further improve the query performance. The experimental results show that the performance of our approach is both efficient and effective, minimizing the response time by decreasing the network transfer cost while increasing the parallelism of I/O and CPU.

1. Introduction

In recent years, with the explosive increase of medical multimedia data in hospital information management systems (HIMS), many applications (e.g., health information retrieval [40], clustering [5,17] and recommendation [39], etc.) require a highly efficient access method to support content-based multimedia retrieval at a large scale. As one of the most important media types, medical images and their management, query, and analysis play a critical role in the modern HIMSs. Millions of medical images generated each year present an enormous challenge to healthcare organizations, as they need to efficiently manage, access, and share such data while trying to reduce costs. How significant is this issue? The following are some eye-opening statistics:

- Medical image archives are increasing by 20–40 percent each year. It is estimated that 1 billion medical images are stored in the US by the end of 2012.
- It is estimated that medical imaging information storage constitutes one-third of the global storage demand, which was the equivalent of 1.2 billion average hard drives in 2007.

* Corresponding author. Address: Apart.502, Unit 2, Blg.7, Qinchunfang, Hangzhou, PR China. Tel./fax: +86 571 87247080.
E-mail address: zy158cn@gmail.com (Y. Zhuang).

0020-0255/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ins.2013.10.013

Although considerable amount of research efforts have been carried out on medical image indexing and similarity query processing in high-dimensional spaces [20], most of these focus on a centralized way (i.e., single- PC-based), which cannot scale up well to large data volume. The query efficiencies of these centralized methods are unsatisfactory because the response time is linearly increasing with the size of the searched file. Therefore the design and development of high performance medical image query methods becomes a critically important research topic.

Cloud services (also known as "the cloud") refers to a network of servers connected by the Internet (or other networks) that enables users to combine and use computing power on an as-needed basis. Individual user does not have to purchase and maintain his own computing power. The cloud provides virtual centralization of applications, storage, etc., which can be accessed by any web-friendly device (such as computer, laptop, smart phone, tablet) virtually anywhere.

A mobile cloud (Mc) can be regarded as a type of flexible computing infrastructure consisting of many computing nodes, which can provide resizable computing capacities to different users anywhere anytime. To fully harness the Mc power, efficient data management is needed to handle huge volumes of medical image data and support a large number of concurrent end users (e.g., physicians). In addition, the Mc environment provides us with location-based query support that enables clinical users to retrieve patient records and images conveniently. To achieve this, scalable, high-throughput, location-based querying, and indexing schemes are generally required. However, as shown in Fig. 1, for Mc-based medical image query processing, exploring parallelism in the Mc to speed up the queries is a new research topic, which has received little attention so far. The challenges include three main aspects:

(1) ***High computation cost in medical image retrieval***: most of medical images are characterized by *high pixel resolution*, *high-dimensional*, and *large-scale*. So the query cost of such medical images is very high.
(2) ***Mobility of Mc users***: most clinical users in the Mc are constantly moving. That means the spatial position of each user varies with the variance of time. So, how to perform an optimal data placement is also a challenging issue.
(3) ***Instability and heterogeneity of the Mc***: the nodes in the Mc are instable, that means, some nodes may be down or connected intermittently to the network. The bandwidth between any two nodes in the Mc may be different according to the variance of time. There is no guarantee that the total response time of each query will be similar.

To address the above challenges, we propose an efficient distributed similarity query processing technique ($MIRC$) in the mobile cloud environment. In particular, the $MIRC$ includes four enabling techniques, namely, *learning-based optimal data placement scheme, content-aware and bandwidth-conscious multi-resolution-based image data replica selection scheme*, *priority-based image block data robust transmission algorithm and learning-based dynamic correlated data caching scheme*. We have implemented the $MIRC$ method and extensive experiments indicate that our approach is specifically suitable for the large high-dimensional image data queries. Without loss of generality, Euclidean Distance is used as the underlying distance function in our approach. The contributions of this paper are summarized as follows:

- We introduce a framework of a efficient and robust large medical image retrieval in the mobile cloud environment ($MIRC$) to improve search efficiency, especially for large-scale high-dimensional image repositories.
- We design a *learning-based adaptive data placement scheme* to maximize the query parallelism at the master node level, in which clinical users' moving trajectories among different departments are sampled and analyzed to estimate optimal placement positions of the medical image data among different departments.
- We propose a *content-aware and bandwidth-conscious multi-resolution-based image data replica selection scheme* to adaptively perform the data transmission process in a reasonable time.

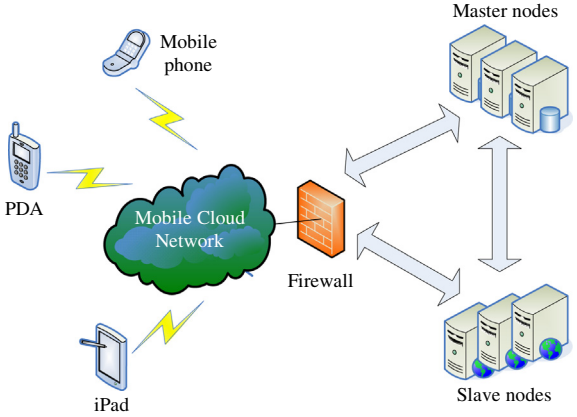

Fig. 1. The general architecture of a mobile cloud environment.