

Danielle Tinio

CIS II

DICOM in Dart (DCMiD) Background Reading

4 April 2014

Separation of metadata and bulkdata to speed DICOM tag morphing

This journal article proposes a new method to reduce the time needed for parsing DICOM data, which is a standard for handling, storing, and transmitting information of medical images such as CT and MRI scans. The authors of this paper emphasize that an increased speed of DICOM metadata transfer is significant for users such as large hospitals which process hundreds of images daily and regional Health Information Exchanges that manage billions of images.

Most medical images are stored and shared based on the DICOM standard, in which pixel data and its associated metadata are stored as Single-Frame DICOM. Each image in DICOM is arranged with four levels: patient, study, series, instance. Because DICOM communication is based on single frame objects that each contains four levels, every time an image is transmitted, there is repetition in parsing and validating metadata. However, there are situations, such as in tag morphing, where only accessing metadata and not pixel data is needed. Tag morphing is the adding, deleting, or modifying attributes of a DICOM study so that it is usable in its target domain. In this case, the standard model is inefficient.

The authors of this paper start with an overview of Multi-frame DICOM (MFD), an already existing method to group all DICOM objects that belong to a series into a single DICOM object. This method removes the need to repeat study and series level attribute for frames in the series. They argue that two drawbacks of MFD are that it has not been

implemented, and there is no standard way to convert older studying into MFD objects. As a result, the authors propose an extension to the MFD method by replacing it with a Multi-series DICOM (MSD). An entire study would be stored in a single MSD object, removing the repeated attributes that occur in the conventional single-frame object.

For the purposes of this paper, the authors define bulkdata to be greater than 256 bytes. The authors' goal was to achieve the following: to aggregate all instances into one data structure, to separate metadata from bulkdata, to eliminate duplicate attributes. In order to access the metadata using the conventional structure, one must parse and validate each single-frame object, which not only contains the metadata, but also the pixel data. By storing metadata and bulkdata separately, the MSD format allows for faster access to metadata by replacing the bulkdata attribute value with a reference to the attribute's location in bulkdata, which is now separately stored. The advantage of MSD is that metadata can be accessed from one file as opposed to accessing multiple SFD files. Below is an image from the paper which demonstrates the current single frame model and the proposed MSD model, which contains one object for metadata values and one object for bulkdata values.

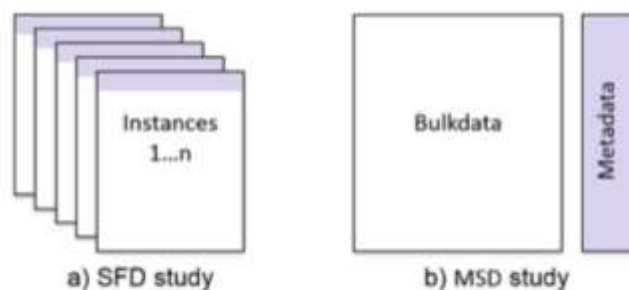


Figure 1: Structure of a Single Frame DICOM (SFD) study versus that of a Multi-Series DICOM (MSD) study.

In order to achieve the need to parse images individually and separate the DICOM data into bulkdata and metadata, the authors created a new value representation, BD (for bulkdata), with the following structure:

VR 2 bytes	Index in BD 4 bytes	Offset 4 bytes	Length 4 bytes
---------------	------------------------	-------------------	-------------------

These 14 bytes then replace what were originally the 256 or more bytes of bulkdata.

The authors use a Java implementation of the MSD format using the dcm4che2 toolkit to explore and compare reading and writing studies in Single-frame DICOM and Multi-Series DICOM. Using six different DICOM studies as input (three MRIs and three CTs), the authors compared the difference in speed between the SFD and MSD methods.

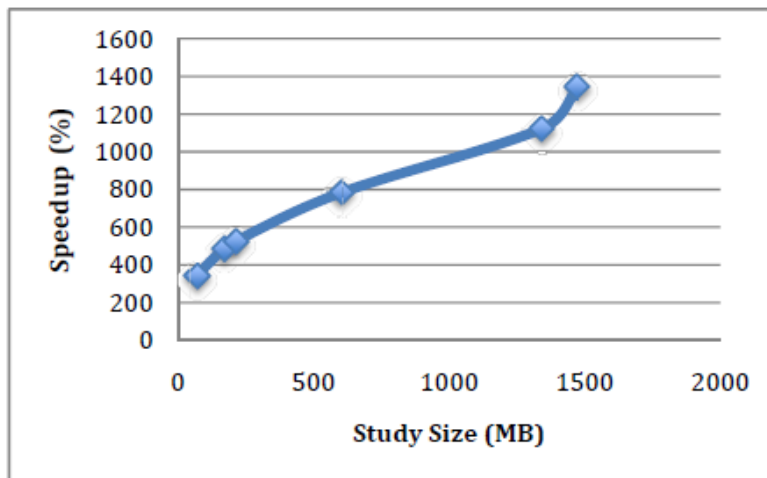


Figure 4: Tag morphing speedup versus study size.

The size DICOM study sizes range between 70 MD and 1.5 GB. The authors' results show that MSD is faster than SFD when accessing metadata, and the difference between the speeds increases as the size of the DICOM study increases.

The article gave a straightforward incentive to explore alternatives to the standard means of accessing metadata, and a clear hypothesis with an easy to follow outline of the

proposed Multi-Series DICOM format and how it performed when compared to Single-frame DICOM. It gave me a more intuitive understanding of our project and why our mentor is invested in decreasing the time needed to access the desired data. Figure 1 clearly demonstrated the difference between SFD and MSD. Initially, it was difficult to understand the motivation behind fast access to metadata. However, the simplicity of the explanations and the flow of the article is something that I want to emulate in future presentations and documentation, particularly for people who are not familiar with DICOM.

This article is specific in that the topic is not as well known in the scope the entire research community and is directed toward a relatively small subset of researchers. Most of the references of this paper are to papers written by the same authors. Additionally, one of those references then only has the DICOM standard as its only reference. The topic is in its infant stages and the interest in faster DICOM parsing is new. However, it is becoming progressively more important to seek ways to improve efficiency to follow the increasing use of medical imaging storage, transmission and handling.

Reference

Mahmoud Ismail ; Yu Ning ; James Philbin;

Separation of metadata and bulkdata to speed DICOM tag morphing

. Proc. SPIE 9039, Medical Imaging 2014: PACS and Imaging Informatics: Next Generation and Innovations, 903905 (March 19, 2014); doi:10.1117/12.2043933.