318-55 38 39

On standards for the storage of images and data

M.J. Haney, R.L. Johnston, and W.D. O'Brien, Jr.

Bioacoustics Research Laboratory, Department of Electrical Engineering
University of Illinois, 1406 W. Green Street, Urbana, Illinois 61801

Abstract

A standardized format is required in order to communicate image information between different systems. It is important that the standard be sufficiently flexible to allow for the encoding of the raw data that was used to form the image. Arguments are presented concerning the importance and value of storing the raw data, independent of its source. Future standards evaluation committees are asked to consider raw data storage flexibility as a selection criterion. An example of image/data storage flexibility in currently used formats are given.

Introduction

Medical imaging systems tend to fall into two classes: systems that produce non-computed images, and systems that produce computed images. Film based radiograms involve no computation. Even when the film is measured and digitized into a matrix of picture elements, the image is still non-computed, since the resulting digital image is nothing more than a numerical representation of the orignal image. As long as the sampling is sufficiently fine (to prevent aliasing), no information is gained or lost as a result of computation. For the same reason, digital fluoroscopy is also a non-computed imaging modality.

On the other hand, a tomographic image is clearly computed. Without the reconstruction process, the raw data is almost meaningless for diagnosis. From the clinical standpoint, the data is valuable only in that it can be used to form an image. This is also true, to a lesser extent, of other imaging systems, such as gamma cameras and ultrasonic imaging systems. In a gamma camera, the raw data are the signals from the different channels. In an ultrasonic B-scan imaging system, the raw data are the received sound signals as a function of time and transducer angle. Processing is required to produce an image (e.g. to calculate position from interchannel comparison, or to adjust the gain to compensate for attenuation).

For practical purposes, raw data is a by-product of the computed imaging process. However, for the sake of generality, one can consider the original non-computed image as raw data. Although most of the following arguments are intended to be applied specifically to the raw data of computed images, non-computed image "data" can be included.

The data obtained for medical imaging tends to be processed in a manner that enhances visually desirable features, for the sake of the viewer. As a result, the original data that contributed to the image are often not recoverable from the image alone. In some sense, a certain degree of reality is lost in the processing.

In most commercial medical equipment, the raw data are not available to the user. However, for those systems (commercial or experimental) which do allow access to the raw data, the merits of preserving the data, as well as the image, should be reviewed. Keep in mind that there is no excuse for storing data simply to prevent a loss of information. Storage space is too valuable to waste in such a manner. Data storage should be considered only when there is a possibility that the data will be used to form new and different, and most important, useful images. Further, if image information is to be exchanged, a great deal of benefit could be obtained by making the format of exchange sufficiently general to handle both images and raw data. With the ongoing development of complete picture archiving and communication systems (PACS) to handle the storage and exchange of images, it is important to consider the added benefits of storing and exchanging data as well. A PACS that can accomodate raw data may be very desirable from the standpoint of the following arguments.

Reasons for storing data

There are several reasons for storing data. Four important reasons are:

For the sake of consistant records

As new data processing methods are developed, it may be desirable to upgrade old records (based on original data) through the new processes. Economically it is unlikely that a hospital could afford to rebuild all of its records every time a new program is created. However, if data is stored, upgraded images could be formed on a case by case basis (on request), using the old data. This has the advantage that as new images are collected, there would be a common basis for progressive diagnosis. Old images (old data under new processing) could be compared directly and validly with new images.

For the sake of evaluating new or competative processing methods

Currently, when a new data processing method is developed, it is necessary to collect data specifically for the purpose of verifying the method, and comparing it to previous methods. It is possible to eliminate this costly developmental overhead through routine raw data storage. If data is stored in the course of normal image generation, special data collection is not required. Moreover, medical images tend to naturally acquired a wealth of auxilliary information as a result of their use for diagnosis. Patient history, results from other tests, and unfortunately, post-mortem examinations, tend to accumulate in time. This information can be useful in verifying new processing methods against actual subject condition. Such detail is rarely available from pilot studies (on human subjects) intended for method comparison. Data storage can provide a valuable benchmark resource.

A similar situation exists for competative processing techniques. There are several different algorithms (gray scale modification, smoothing, edge sharpening, ...) for enhancing images to accentuate different features. In many cases there are also different algorithms for forming the images in the first place. (For tomography, there are algebraic reconstruction, convolution - backprojection, direct fourier space methods, and many more.) Storing data allows alternate methods to be applied later if some question arises concerning the accuracy of the original image processing. Not only can alternate analyses be applied to the images, but also alternate images can be obtained from the data. This is especially important with imaging modalities that tend to produce relatively large amounts of processing generated artifacts (e.g. divergent beam tomography).

For the sake of communication

There are two important reasons for transferring images from one place to another. First, by allowing a patient's records (images included) to follow the patient from hospital to hospital, it is possible to maintain an accurate, up to date record of the patient without redudant examinations (which can be costly, and in some cases, hazardous). A second reason is to allow images collected at one site to be processed and interpreted at another, better equiped site. This is the basis of teleradiology.

However, for images to be transfered, some form of information communication standard must be used. When information is transfered on a physical media (i.e. magnetic tape, floppy disk, or video disk; as opposed to transmitted by radio or over wire), the communication standard is also a storage standard. Common sense suggests that if the same standard is used for both communication and general storage (archiving), then a minimum of special purpose software will be required. The same programs can equally well be applied to old (stored) as to new (communicated) data.

If the same communication standard can support the transmission of data as well as images, an important possibility arises. Standardized image and data transmission opens the door for low overhead medical image research at the academic level. For the cost of computation, a small college or university could begin a research program in medical imaging, using data routinely collected by larger institutions (e.g. medical centers, hospitals, research laboratories). This would eliminate the major capital expenditures required to obtain imaging equipment. (Note: such expenditures are typically passed on to government or industry, and no one realy profits.) Also, by bringing medical imaging opportunities to the classroom, better prepared students will become available for both industrial and clinical work.

For the sake of non-image processing

It is virtually impossible to measure tissue properties from enhanced images. Too much real information has been lost. Quantitative analyses require uncontaminated input. The same is true of tissue characterization (for automated pattern recognition). Raw data are required in order to make valid comparative measurements, or decisions based on fundamental characteristic properties. Even if such analyses are not done currently, routine data collection allows for an outside agency to perform the processing (for its own research purposes, but at its own cost).

Unfortunately, there are also several reasons for not storing data. The most important is cost. Considering the amount of space and storage media required, data should not be stored if it is not going to be used. The same is true of medical images. Nonetheless, short term data storage (weeks, monthes) has most of the benefits above, with minimal costs. An additional cost advantage is available by storing the data in the same format as the image.

Data storage requirements

Images are conventionally two dimensional: I(x,y) where both x and y are measures of distance. Often, in the collection of tomographic images, a third dimension (again distance) is introduced: I(x,y,z). When time-series images are studied, the third dimension is time rather than distance: I(x,y,t). Time-series volume measurements are also conceivable: I(x,y,z,t).

In some cases, images are produced at different energy levels, or frequencies, or modes (time of flight, attenuation), etc. These multimode imaging schemes add dimensions of their own: I(x,y,m1,m2,...). Thus medical images can be thought of as arrays of numbers, where the number of dimensions of the array depends on the type of imaging system.

Raw data are also multidimensional by nature. Measurements are made of intensity, energy, frequency, temperature, etc., as a function of time, temperature, position, angle, etc. The number of dimensions involved, as well as the units of measurement, depend on the specific type of examination being performed. However, in the same sense as the images described above, data can be readily stored as a multidimensional array.

Two examples in our laboratory include the ultrasound computed tomography (UCAT) system and the scanning laser acoustic microscope (SLAM), both of which are interfaced to the laboratory computer (Perkin-Elmer 7/32) for data acquisition, analysis, and storage. For the UCAT system, the raw data which requires storage, depending on the particular experiment being conducted, include the transmitted beam vector (x1,y1,z1), the received beam vector (x2,y2,z2), time (t), frequency (f) in the case of narrow band operation (or time reference (t1) of the onset of the pulse in the case of broadband operations), temperature (T), and selected system variables to indicate the operating mode and ultrasonic output power. The SLAM raw data include the specimen position (x1,y1,z1), laser beam position (x2,y2,z2), time (t), frequency (f), temperature (T), and selected system variables.

Commonly used image formats reflect the multidimensional nature of the image information. For example, the National Cancer Institute (NCI) image/data format (as used at the Mayo Clinic[1]) uses three dimensions: view angle, line, and sample. Each data record is a vector of samples for one line or view angle. The record includes two indices that determine the line number or angle. A header record precedes the data records, and specifies the number of views, number of lines, and number of samples per line, along with several other descriptive parameters. A conventional image is represented by filling each data record with one raster line of values, and incrementing the line index for the next record. (Only one view angle is used. If several images are to be stored in one file, they can be assigned to different view angles.) Data can also be stored in the NCI format by taking advantage of the format's three dimensional structure. Tomographic data, for example, can be stored using the data record to hold the projection (line integral) data, and indexing on the view angle.

Unfortunately, the NCI format is limited in its ability to store higher dimensional images or data. (This is understandable, since it was not designed to be a universal image format). In order for an image storage standard to be truely general, it must be able to encode a multidimensional array of values, for a variable number of dimensions. Since raw data is also an array of values, it follows that any multidimensional image storage standard is automatically a data storage standard as well.

318-53

Conclusions

There are three principle conclusions to be drawn. First, the storage of raw data (as well as images) can benefit both the institution forming the images and other research groups. Second, standardized forms of information representation are required for the exchange of images. Third, any sufficiently powerful (multidimensional) image standard can also be used as a data standard.

The generation of such a standard remains. Standards tend to spring either from a single group with overwhelming clout, or from committees created to study, deliberate, and eventually decide the issue. Whatever its source, it is asked that multidimensional information storage capability be considered as an important criterion for choosing an image/data storage standard.

Acknowledgements

This work is supported in part by the National Institute of Health's National Institute of General Medicine Science (GM 24994) and National Cancer Institute (CA 09067), the National Science Foundation (ENG 76-22450), and the University of Illinois' Bioengineering Program.

References

1. Rob, R.A. Construction of a Projection Data Base for Testing Algorithms for Computerized Transaxial Reconstruction, report perpared for the National Cancer Institute (Contract N.C.I. - CB - 53857 - 33) by the Mayo Foundation, 1977.