Data creates more data : A study of the Datasets used by Virtual Heart Research Team at University of Malaya

Fariza Hanum Md Nasaruddin*

1 Faculty of Computer Science and Information Technology, University of Malaya, Lembah Pantai, 50603, Kuala Lumpur, Malaysia
Phone: (60)-12-2105955; Fax: (60)-3-79579249
Email: fariza@um.edu.my

Data is an important component in any research. Data can be newly acquired at the initial stage of a research project or it can be an already available data which has been stored somewhere. In the case of the research conducted by a research team at University of Malaya, the datasets used are cardiac data. These datasets will be processed and most often will yield another new set of data. In this paper, datasets used by a group of researchers in the area of virtual heart modelling in the Faculty of Computer Science & Information Technology, University of Malaya, were studied to understand the transformation of the datasets from one form into another. This paper will illustrate the different forms and formats of the datasets. For example, a dataset used by this group can transform from a simple set of ASCII data into a very complex mathematical model. On the other hand, an image dataset can be processed to produce a set of textual or numeric data in ASCII form. The final objective of this study is to get a set of requirements to assist in the development of a data repository for the research group. Because of the multiple forms and formats of the data involved, the design and development of the repository will be a challenging task.

1. Introduction
Heart disease has been one of the leading cause of death in the world for many years now. According to the World Health Organization (WHO), each year, an estimated 17 million people die of heart diseases, including coronary heart diseases and strokes. This number however, is decreasing in the developed countries, as a result of medical advancements and knowledge gained from researches conducted in this area. These researches may be clinical researches with the purpose of improving the diagnosis for such diseases and improving cardiac care after diagnosis. Clinical researches make use of clinical or medical data to determine and understand the clinical and medical aspect of cardiac behaviour. For example, to understand the risk factors and how they effect different population, or determine the reaction and the response of hearts to certain drugs, to determine how to best measure the performance of the heart, etc. Some of these researches are invasive while others are non-invasive in nature. Some researches are safe enough to be conducted on human patients while others can only be carried out on animal hearts due to the risks involved. Thus, researchers are always looking for new tools to assist them in carrying out these tests. With better power and multimedia abilities, computers are now being used to model a human heart. These new type of researches are non-clinical and are conducted by researchers in areas such as bio-medical engineering, computer science, and information technology. As with clinical research, data is also an important component in a non-clinical research. Data can be newly acquired at the initial stage of a research project or it can be an already available data which has been stored somewhere. Although there are many data sources available for many research interests, data for cardiac dynamics and data to be used in a non-clinical heart research are quite limited, especially in Malaysia. The problem of lack of data is apparent in the case of a project team in Faculty of Computer Science and Information Technology (FCSIT), University of Malaya (UM), Kuala Lumpur.

2. Virtual Heart Project at FCSIT, UM
The heart is a very complex organ, and thus it is not easy to model. For that reason, many researchers in this area try to focus only on certain aspects of the heart in their research. Over the years, researchers have come up with various models that either mimic certain function(s) being carried out by certain parts and/or mechanisms available inside the heart, or the function of the heart itself as a whole, usually seen from various ways that range from the physical aspects such as the blood flow [3], apparent pressure changes in the beating of the heart [4], and mechanics and rhythm to biomechanical, such as the behavior of the aortic valves to topics such as electric fields and currents [1, 11].

In University Malaya (UM), a research group from Faculty of Computer Science and Information Technology (FCSIT) has been working on a virtual heart project as part of a bigger project of a virtual human. The virtual heart project concentrates on the heart wall mechanics analysis and fluid mechanics. The main interest of this research team is to build computer applications and computerized models of the heart with the intention of making them realistic and true enough to be used by cardiologists for a variety of purposes.

The team was working on a number of projects and has successfully produced five masters thesis, with a number of PhD thesis in progress. The team started their project with the development of a simple application program to measure key
cardiac parameters from Magnetic Resonance Images (MRI) of the left ventricle of a heart.[8]. In this project, the researcher extracts the boundaries of the heart’s left ventricle from 2D MRI datasets in order to measure them to calculate the relevant cardiac parameters. This was done by employing techniques for segmentation and contouring, and in this case, the active shape modeling and active contour modeling techniques were used. Another approach was taken in another project, where the segmentation of the same 2D MRI datasets were produced using a Pulse-Coupled Neural Network [7].

Another researcher built a dynamic cardiac mechanics model based on Fiber-fluid model using the same MRI data. While there are a number of other projects, this paper will only discuss the data used in these three cases to illustrate that data will create new sets of data which can be used by other researchers. In addition to the three cases, this paper will also discuss two more cases which will illustrate that there are also data in other formats such as the ASCII file format and point cloud format, and these data can also be transformed into image data and vice versa.

2.1 Case Study 1

In the early stages of the research, the group was able to acquire some MRI data of a healthy human volunteer from University Malaya Medical Centre. By using appropriate software, the MR images were converted to Tagged Image File Format (TIFF) format. Sixty-seven (67) image slides were prepared from the original MR images. It was a tedious task. MR images had to be read from a proprietary format on a Magneto Optical Disk (MDO). Tools such as Materialise Mimics was used in this process. Before any work can be done on these data, they need to be “cleaned” first. These data need to be pre-processed. For example, noises from the original image need to be eliminated. In the case of these images, an anisotropic filter that reduces image noise was applied to them.

The MR image also covers a wide area of the volunteer’s body, much of which is not related to the heart. The area of the heart which is the focal point of the research is known as the region-or-interest (ROI). Such, the ROI of the image need to be extracted. The process of extracting the ROI is called cropping.

The images were cropped and a specific segmentation technique was used to extract the boundaries of the heart’s left ventricle (LV). The boundaries of the LV were then used as data in the software application to calculate the relevant cardiac parameters.

Figure 1 illustrates the transformation of the original data into other forms of data [8].

2.2 Case Study 2

In the first case, the MRI image was segmented using the active countour model (ACM) approach. However, this technique is not without its drawbacks. In this particular technique, an initial curve model is formed and then internal and external forces will act on this model to find the boundaries of the region of interest [9]. The original snakes algorithm has its problems such as the initial curve must not be too far away from the edges of the target object it is trying to detect. Thus, another researcher tried to find other techniques which could be used to segment the MR images. In this second case, a researcher used an artificial intelligence technique called the Pulse-Coupled Neural Network technique to segment the same MR images used by the first researcher.

Looking back at the first case, the project started with the availability of raw MRI data. The second researcher in the second case can choose to process the raw MRI data into a new set of TIFF files. If she chose to do that, she will also have to crop the image in order to isolate the ROI to minimize any artifacts in the image. However, since this has been done by the first researcher, she chose to start her research by selecting the cropped images, and by doing so, save many hours of repeating the same tasks which had been carried out earlier by the first researcher. She was able to skip the following tasks:

- convert the raw MR Images into frames (TIFF files)
- crop the image into region-of-interest (ROI)

However, when she tried to process the image using a pulse-coupled neural network (PCNN) [5], she found that the images taken from the previous researcher were “noisy”, so she had to employ certain techniques to reduce the noise. In doing so, she had created a new set of data. She then proceeded to use this new set of data, and by using PCNN, managed to produce another set of segmented data.
The second researcher now has produced 2 new sets of data from the original MR Images as illustrated in Figure 2 [7]. It is important to note that in the two cases discussed above, the same cropped images were used by two different researchers to study two different approaches to image segmentation. Because the original images were the same, the result of the two different segmentation techniques can easily be compared.

2.3 Case Study 3

The earlier two cases illustrated the creation of new sets of image-based data from the MR images. In this third case, we will see that these image-based data can give birth to a three dimensional (3D) model. In this particular research, the researcher was interested in implementing the fiber-fluid model formulated by McQueen and Peskin [6]. To do that, the researcher needed a 3D model of the heart. Due to limited data available, he built his own 3D model from the 2D images which are available at the faculty [10]. Using the same cropped and cleaned image taken from the same MR images, an edge detection algorithm was used to identify and extract the ROI from the images. All 67 frames from the MR image were processed and the contour of the heart LV from each frame were obtained. The MR machine captured the image from several layers of the heart. Each of the layer is referred to as a slice. These slices are then stacked as seen in figure 3, and a whole heart was reconstructed.

The stack of the contours were scaled, transformed, and then interpolated to produce a beautiful 3D model of the heart’s LV. Using this model, the researcher then proceed to simulate the fiber-fluid model using specific mathematical equations. So in this example, we see that from the same MR images, a series of 2D cropped ROI gave way to a 3D model. The process taken to produce the 3D model involved various activities and were tedious. Again, it would be very helpful when such a model can be stored so that it can be shared with another researcher who needs it.

3. Data creates more data

The cases discussed earlier illustrates only a fraction of the possibilities of data which can be created from one set of raw data. The projects had shown that from one set of MR images, other datasets were created. The MR images were not the only “raw” data used by the research team at FCSIT, UM. In another project, a different set of data was used. Here, a heart was scanned using a 3D scanner. Again, this dataset need to be manipulated and processed further before any reconstruction work can take place.

![Fig. 2. Segmentation using PCNN](image)

![Fig. 3. From 2D images to 3D model](image)
mathematical model on a finite element platform. Figure 4 illustrates [2].

4. Conclusion and Future work

The study of the datasets being used in FCSIT, UM has given us a glimpse of how one dataset can create more datasets. Examples discussed above illustrates that raw data need to be processed before it can be used in any particular projects. Available MR images cannot be directly used unless they have been processed into digitized versions. Data from scanned heart too need to be digitized, and processed further before they can be used to reconstruct a model. The effort needed to transform data from its original state into formats that can be used in virtual heart projects is quite substantial. Therefore, it would make a lot of sense if the new sets of data which are born from the original data are stored in a repository. The stored data can be made available to other potential researchers. By doing so, researchers starting on a new project can save a lot of time and not waste any effort in finding new datasets, preprocessing them, and cleaning them up. Researchers can even choose to use an already reconstructed model rather than reconstructing the same or similar model all over again.