International Journal of Computer Applications (0975 - 8887)
Volume 176 - No.9, October 2017

Role of Software Engineering in Visualizing Large Volume of Hyperspectral and Medical Data Sets

Basaeir Y. Ahmed
Computer Science Dept.
Science College
Basrah University

Safa A. Najim
Computer Information System Dept
College of Comp. Sci. and IT
Basrah University

Widad A. Mansour
Biology Dept.
Science College
Basrah University

ABSTRACT

The software engineering is a domain who cares for the production of the software development with high quality in response to the requirements of the market and delivered on time. This paper will discuss the software engineering, in their applications, and software development to deal with big data, as hyperspectral and medical imagery data sets. The role of the software visualization in the interpretation of the data, where the results will be presented to the user's clearly and beautiful. A color space is a method to specify, create and visualize color. There are several types of systems colors for example RGB, CIE XYZ, HSV and HSI. Color mapping has an important role in the process of the visualization to understand data.

Keywords

Computer Science, Software Engineering, Information Visualization, Medical Imagery

1. INTRODUCTION

Software engineering is branch knowledge interested in the software production with high quality, less expensive and handed in due course. Software is not just a program itself but also all documents related, configuration data required to make this software is working correctly[29]. The system usually consists of separate programs, and configuration files that are used to set up these programs. System documentation is included to describe the structure of the system. Explain how to use the system, and websites for users to download the latest product information . The goal of software engineering is to create appropriately work and build high quality programs [19].
Large data sets, as hyperspactral and medical images, contains hundreds of samples per pixel [17]. Visualization is the central part of the discovery and analysis of data [14]. The visualization is a simple way to understand the high dimension because the relationships between the original data are ambiguous [20]. One of the general purposes of the multidimensional images visualization is data analysis of hyperspactral [11]. However, it is a challenge to display all useful information in such visualization. The Common methods are used a representation color of red, green and blue (RGB) to provide a quick glimpse at the scene [27]. RSI contain a large number of bands and cannot be displayed directly in the color scheme for only three layers. RSI initially need to be converted to low dimensions to display the visualization while preserving the maximum amount of information [2].

2. SOFTWARE ENGINEERING

Software engineering is engineering discipline that is concerned with all aspects of software production from the early specifications of the system . Software engineers focus on software development produced (specified by the customer). They are of two types: generic products and customized product. In generic products, these software type can be used by any one. Examples drawing packages, word processors, and databases. Customized products are used for the particular and especial user [29] [1]. Examples traffic control systems, control systems for electronic devices, and systems written to support a particular business process.
A good software must take into account to four qualities [29]. Maintainability software, is the first quality, which must be written in programs that can evolve to meet the changing needs to customers. This feature a crucial software development and important condition for variable business environment. The second is the dependability and security, where the reliability program includes set of features including the reliability, security and safety. Software reliable does not cause physical damage or economic in the case of the failure of the system. Third, efficiency software should not extravagance in the use of the system resources such as memory and processor. Includes the efficiency of the response, the processing time, and memory usage and others. The last, acceptability software must fact that the acceptable to users of the system designed from them. This means that it must be understood and use and compatible with other systems which they use.
A software process is a series of activities which lead to the production of software [29]. In order to work effectively :-

(1) Software specification: includes the customer and software engineers to define the required software.
(2) Software development: the codes of software are written.
(3) Software validation: which are checked the program to make sure it's match specifications requested by the customer.
(4) Software evolution: where are modified programs to reflect the requirements for the customers and market changing.

2.1 Software engineering and raw data

The software engineering is important in many applications, where it is able to product reliable systems. Which are trustworthy eco-

nomically and quickly [29][9]. In addition, applying software engineering technicians generated cheaper software. Software engineering permits to do change develop and maintain on the software in any easy way to keep up the development.

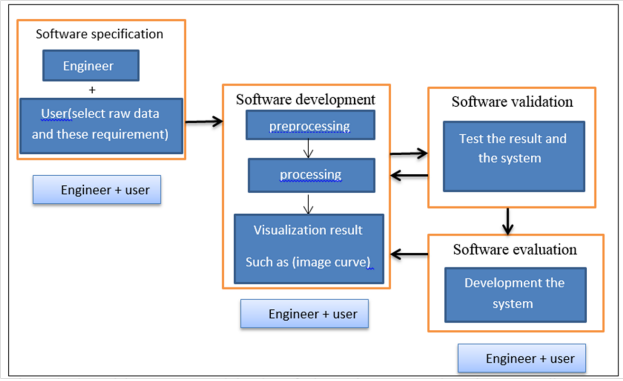

Fig. 1. *Models to process and develop of a huge data system where the user or client's always be found in stages of the system with the possibility of reception requests at any time of stages of the system.*

Raw data are the data that contain information which is very important and many fields. Such as data hyperspactral, photos, medical and simulation physical processes various. In order to run this data, we should do the following features [29]:

(1) Acquisition: Where this data is? and How can we acquire it?
(2) Preservation: what are the requirements of storage for data?
(3) Preprocessing: Do we need to make any process on the data?
(4) Processing: what are the requirements for processing this data?
(5) Generating / presenting results: what type the results are looking for? and What is their accuracy?

The understanding of the requirements in large data necessary to complete successful project, and therefore it is important to include some principles of engineering requirements in the processing. This allows the team development to be more focused on the requirements and also use the nature of iterative of processing to refine the requirements for the life cycle of projects [29] [9]. As in the Fig.1, which illustrates an option to admission requirements of the customer at any stage of the design of the system without any obstacles.

3. DATA VISUALIZATION

Data visualization aims to contact data clearly and effective through graphical representation. Visualization technique discover and keep the relations between the points which cannot be observed form the original [10][32]. the visualization deals with the data presented to the user is two or three dimensions. Both 'information visualization' and 'software engineering visualization' are data collection activities but the motives are clearly different. Software engineering visualization connects the initial design modules and identifies the connections of the important structures. While information visualization reveals the incomprehensible data structure, which is an important part of the software development process. In particular, 'software visualization' executes the codes to display the incomprehensible and hidden data [25].

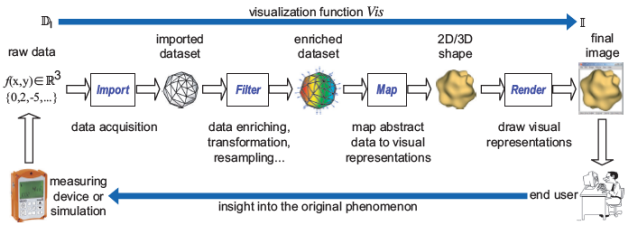

Fig. 2. *Visualization pipeline is through importing where data conversion raw data to the representation of graph can deal with it. Then by filtering is determined important information and improving the data. The step mapping is transforming the data to the data visible, and is viewing information clearly in step rendering. The most important part is a study of the data generated and compare with the data original to answer questions specific this process, come on insight a reverse to visualization [32].*

The purpose of the visualization is to get the insight, as in the Fig.2, which shows the general idea of the visualization pipeline. In computer and information science, the visualization is a visual representation of the domain space using the graphics images sequence of animation, to provide data structure and behavior dynamic complex data representing systems, events, processes ,objects, and concepts [32][33].
Visualization is a way to display the information to allow people to see the unseen data to help in the understanding of the show and exploration or analysis [25][5]. It helps people to extract the accurate information [3], create something beautiful [28]. Color per pixel in visualization is a summary of the information in the data original that have been mapped in projection space. Visualization can be used in several fields, such as computer science, mathematics, cognitive and perception science, and engineering [19].

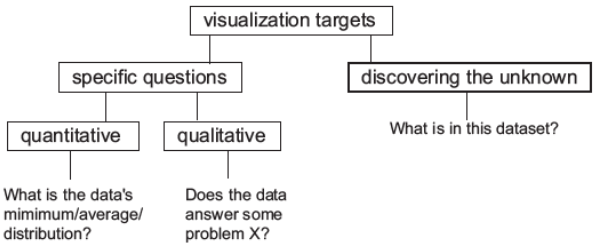

Fig. 3. *The objectives of the visualization. The first goal is to answer specific questions, which may be quantities questions for example, the largest value, less the value of the amended or the qualitative questions for example, asking for a particular problem. The second goal is to discovers information which is not known nature of the data [32].*

3.1 The visualization target

The visualization can answer several kinds of questions. In the subject of insight described two types of information [32] as in the Fig.3: Answer questions about a particular problem, and Facts about a particular problem we do not know what the problem.

The first case focus about the specific question. There are some questions about the specific phenomenon, process or data set. The purpose of visualization is to answer these questions quickly. These questions may be quantitive or qualitative. In quantitative questions, the answer will not be number, but usually set numbers for example, what are the value of the function f (x) values x . We can answer by print values but the best answer is in the graph values. Because the better understanding of the distribution of associated data at visualization. The task is best by using visual representations. In the other hand, the example of the qualitative questions is a picture the nature of a patient certain. The doctor only able to discover the problems The decision-making will be supported by the expert through interactive visualizations.
The visualization able to discover the unknown is the second case of the visualization ability. The visualization may be important to answer the question "what is in the data sets". It can discover the unknown information.
The visualization is divided in the two types:

(1) Scientific visualizations can be described imagine phenomena is three-dimensional such as meteorological, architectural, medical, biological, etc.), as in figure(1.3) that shows the visualization average temperature data rainfall, where the areas of the temperature of high be painted red and regions with a low temperature are shown in the green and fonts shaded represent rainfall [32] .
Fig.4: information visualization average temperatures rain is represented temperatures during the color. Red color represent high temperature and Green color represent low temperature and lines shaded represents the rates of rainfall. Arrow indicates area high rate rainfall and moderate temperature.

(2) Information visualization focus on the objectives of graphs that focus on the user [16]. Data are separate called information visualization. These data, include tables, time series, documents, and software source. As in figure (1.4), it represents the immigration population as areas are represented the points of red. Cities that are among the migration of population large linked lines black-and are grouped lines together to become the flow of the population is visible and this feature is not clarified only through the visualization [32] [22].

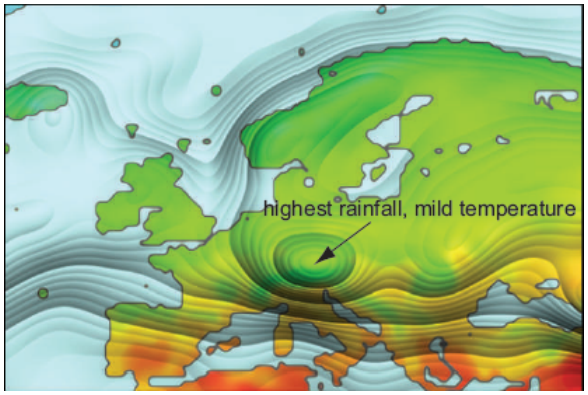

Fig. 4. *information visualization average temperatures rain is represented temperatures during the color. Red color represent high temperature and Green color represent low temperature and lines shaded represents the rates of rainfall. Arrow indicates area high rate rainfall and moderate temperature [32].*

Fig.5: the idea of information visualization can be used with the migration of population. The visualization has cities which are represented by red points. Lase links are used to group cities, where there are different length of that links.

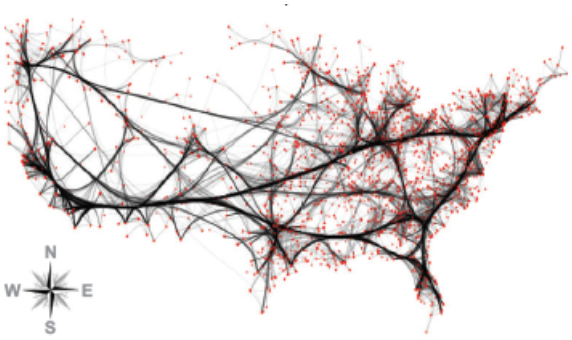

Fig. 5. *the idea of information visualization can be used with the migration of population. The visualization has cities which are represented by red points. Lase links are used to group cities, where there are different length of that links [32].*

4. HYPERSPACTRAL IMAGERY

Hyperspectral Imagery is the art, science, and technology of observing an object, scene, or phenomenon by instrument-based techniques. Earth monitoring collects biological information, geometric, physical and chemical properties around the earth. It helps to evaluate the situation and monitor changes in the natural and cultural environment, give us geospatial data [13] [24].

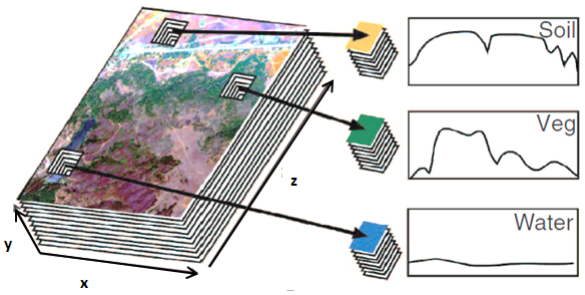

Fig. 6. *Obtain a RSI where the radiation is released to the surface of the Earth with multiple wavelengths and measure the reflection of satellite radiation from the earth thus, there are very large number of bands up to more than 200 Bands which help to distinguish the water, vegetation, and soil by the curve of data reflection [12].*

SRI contains tens or hundreds of narrow adjacent bands and due to the massive data volume, they cannot be displayed these data easily, as in Fig.1. There are several constraints to dealing with RSI such as: [13] [15] [21].

(1) High Cost: The computational costs of processing are very large because it con-tains a large number of bands.

(2) Storage requirements: A single RSI data set contains a large number of bands per scene. The space required to store several hundred megabytes of memory.

(3) Redundancy: The nearby image bands in the RSI carry a high degree of spatial correlation. Due to the continuous and narrow nature of RS sensors . The differ-ence in spectral information between the two adjacent wavelengths is usually very small so, the resulting grayscale bands appear almost identical. There are many data in the scene, therefore it seems to be redundant, but embedded is the important information that can often be used to determine surface material.

(4) Hyperspctral imagery connected to a thin layer of the Earth's surface.

5. MEDICAL IMAGES

Medical imaging techniques are used to discover and understand the diseases development in human body. in order to make diagnostic decision about that. The management of large collection of medical images is a challenge in the healthy centers and hospitals [23]. Because the image database has been rising during the last year, the useful information is required to help the specialist to evaluate the patient status [5]. For example, the most often method uses by a doctor is to look into images slices (slice by slice) to understand the disease development, and sometimes, it is difficult to have decision about that. In addition, if the disease is discover, it is difficult to know the disease severity or progression (early, medium or advance) [31]. Thus, visualizing medical images is a good way to generate accurate image summarizes the main information of all slices and give much better results. Color visualization by dimension reduction method is used recently to reduce a series of medical image slices into a right image [34] [26] [4]. Interpreting and analyzing the color image are easier than original slices [30]. This method tries to explore the right information to user by recognizing the similarities and differences in visual image. The user requires this summarized colour image, which represents the overview of all images, in order to start his process.

Each image in the data sets does not necessary have full knowledge, and sometimes, overall images do not help to explain the patient status. Feature dedication by dimension reduction methods has been used recently because it is successfulness in information visualization [35] [6] [8] [7]. Dimension reduction methods are used to reduce the dimensionality of original data sets into target dimension, where the hidden information in original medical images is revealed in the visual space. In visualization, the dimension reduction method projects the data sets into two-dimensional space (to generate gray image) or three-dimensional space (to generate color image). However, the visualization by dimensionality reduction is hard problem, and its efficiency depends on the used method.

6. VISUALIZING LARGE VOLUME DATA SETS

The steps of the pipelines are:-

(1) Importing Data: In this step, the data is imported. This mean, the representation of the data sets (D_I) will be found. This step is one-to-one mapping, thus, the input data D_I in converted to D to be suitable for visualization pipeline, as in equ.1

$$Import : D_I \rightarrow D \quad (1)$$

$$where\ D \in D_I \quad (2)$$

It should be note, this step is very important because the quality of the final visualization will be dependent of it. The importing data step should try to preserve as much of the available the information of the hyperspactral imagery data sets.

(2) Data Filtering and Enrichment: after imported the data, we must decide what is important information and features would be extract raw data sets into most appropriate. This process lead important filtering data to extract the important information, as in equ.3. The filtered data are important to generate the visualization [33] [16].

$$Filter : D \rightarrow D \quad (3)$$

(3) Mapping Data: The enriched data sets are produced in the filtering step, which should represent the interested features. This step wants to map filtered data sets D into visual domain D_V, and the visual domain is 2D or 3D.

$$Map : D \rightarrow D_v \quad (4)$$

In equ.4, D belong to the data and maps makes the data visible D_V in this step is to reduce the size of the data [33] [18].

(4) Rendering data: this is the last step of the visualization data sets. Rendering takes the 3D representation which was create the mapping operation. The user determines the parameters for displaying for example show points or lighting to set up the required visualization, as in equ. 5.

$$Render : D_v \rightarrow Visualization \quad (5)$$

7. CONCLUSION

Software Engineering is very important technique to build and improve the software. Visualizing large volume data sets, as hyperspctral and medical imagery, is very important in order to see the unseen information. The visualization for this large volumin data sets should find a result, which is suitable for human eye. Preserving information of the orignal data sets is very important to find effecient visualization. This paper explain how the final visualization of data sets represents the simple representaion of the complex structure of the orignal data sets.

8. REFERENCES

[1] A. Abran and J. W. Moore. *Guide to the Software Engineering Body of Knowledge*. United States of America, 2004.

[2] L. Wang D. Liu and J.A. Benediktsson. An interactive color visualization method with multi-image fusion for hyperspectral imagery. *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 1088–1091, 2015.

[3] A. Bezerianos E. Dimara and P. Dragicevic. The attraction effect in information visualization. *IEEE. Transactions on Visualization and Computer Graphics*, 23(1):471–480, 2017.

[4] Samuel Gerber, Tolga Tasdizen, P. Thomas Fletcher, Sarang Joshi, Ross Whitaker, and the Alzheimers Disease Neuroimaging Initiative (ADNI). Manifold modeling for brain population analysis. *Medical Image Analysis*, 14:643–653, 2010.

[5] Patric Hagmann, Lisa Jonasson, Philippe Maeder, Jean philippe Thiran, Van J. Wedeen, and Reto Meuli. Understanding diffusion mr imaging techniques: From scalar diffusion-weighted imaging to diffusion tensor imaging and beyond. *Radio Graphic*, 26:205–223, 2006.

[6] Ghassan Hamarneh, Senior Member, Chris McIntosh, and Mark S. Drew. Perception-based visualization of manifold-valued medical images using distance-preserving dimensionality reduction. *IEEE Transactions on Medical Imaging.*, 30:1314–1327, 2011.

[7] Jihun Hamma, Dong Hye Ye, Ragini Verma, and Christos Davatziko. Gram: A framework for geodesic registration on anatomical manifolds. *Medical Image Analysis*, 14:633–642, 2010.

[8] Zhanli Hu, Jing Zou, Jianbao Gui, Junyan Rong, Yanming Li, Dongxing Xi, and Hairong Zheng. Real-time visualization and interaction of threedimensional human ct images. *Computers*, 5:1335–1342, 2010.

[9] B. Hollein J. Al-Jaroodi and N. Mohamed. Applying software engineering processes for big data analytics applications development. In *Computing and Communication Workshop and Conference (CCWC), 2017 IEEE 7th Annual, Las Vegas, NV, USA*, 2017.

[10] M. Kamber J. Han and J. Pei. *Data Mining Concepts and Techniques*. 2012.

[11] M. Kazhdan D.Lepage J. Lawrence, S.Arietta and C. OHagan. A user-assisted approach to visualizing multidimensional images. *IEEE Transactions On Visualization And Computer Graphics*, 17(10):1487–1498, 2011.

[12] G. Camp s-Valls P. Scheunders Nr M. Nasrabadi J. M. Bioucas-Dias, A. Plaza and Jocelyn Chanussot. Hyperspectral remote sensing data analysis and future challenges. *IEEE Geoscience and remote sensing magazine*, 1(2):6–36, 2013.

[13] G. C. Huurneman K. Tempfli, N. Kerle and L. L. F. Janssen. Principles of remote sensing. In *The International Institute for Geo-Information Science and Earth Observation (ITC), Netherland*, 2009.

[14] S. Kaski and J. Peltonen. Dimensionality reduction for data visualization. *IEEE Signal Processing Magazine*, 28(2):100–104, 2011.

[15] Ketan Kotwal and Subhasis Chaudhuri. Visualization of hyperspectral images using bilateral filtering. *IEEE Transaction on Geoscience and Remote Sensing*, 48, 2010.

[16] G. Wang L. Wang and C. A. Alexander. Big data and visualization: Methods, challenges and technology progress. *Digital Technologies*, 1(4):33–38, 2015.

[17] J. Hu M. Cui, A. Razdan and P. Wonka. Interactive hyperspectral image visualization using convex optimization. *IEEE Transactions On Geoscience And Remote Sensing*, 47(6):1673–1684, 2009.

[18] K. Ma and S. Parker. Massively parallel software rendering for visualizing large scale data sets. *IEEE Computer Graphics and Applications*, 21(4):72–83, 2001.

[19] N. M. Munassar and A. Govardhan. A comparison between five models of software engineering. *IJCSI International Journal of Computer Science Issues*, 7(5):94–101, 2010.

[20] Safa A. Najim. Information visualization by dimensionality reduction: a review. *Journal of Advanced Computer Science and Technology*, 3(2):101–112, 2014.

[21] Safa A Najim, Ik S Lim, P Wittek, and M Jones. Fspe: Visualization of hyperspectral imagery using faithful stochastic proximity embedding. *IEEE Geoscience And Remote Sensing Letters*, 12(1):18–22, 2015.

[22] Safa A. Najim and Ik Soo Lim. Trustworthy dimension reduction for visualization different data sets. *Inormation Science*, 278:206–220, 2014.

[23] Safa A. Najim and Widad A. Mansour. Hybrid visualization of the medical images data sets. *International Journal of Computer Application*, 136(8), 2016.

[24] Safa A. Najim, Alaa A. Najim, IS Lim, and M Saeed. Parallel faithful dimensionality reduction to enhance the visualization of remote sensing imagery. *Neurocomputing*, 168:560–565, 2015.

[25] F. T. Marchese O. C.Z. Gotel and S. J. Morris. The potential for synergy between information visualization and software en-gineering visualization. In *12th International Conference Information Visualization IEEE, London, UK*, 2008.

[26] Aljabar P., Wolz R., Srinivasan L., Counsell S. J., Rutherford M. A., Edwards A. D., Hajnal J. V., and Rueckert D. A combined manifold learning analysis of shape and appearance to characterize neonatal brain development. *IEEE Transaction on Medical Imaging*, 30:2072–2086, 2011.

[27] S. Cai Q. Du, N. Raksuntorn and R. J. Moorhead. Color display for hyperspectral imagery. *IEEE Transactions On Geoscience And Remote Sensing*, 46(6):1858–1866, 2008.

[28] F Mao S Grainger and W Buytaert. Environmental data visualisation for non-scientific contexts: Literature review and design framewo. *Environmental Modelling & Software*, 85:299–318, 2016.

[29] Ian Sommervill. *Software Engineering*. Addison-Wesley, 2011.

[30] Richard Souvenir and Robert Pless. Image distance functions for manifold learning. *Image and Vision Computing*, 25:365–373, 2007.

[31] Adam Switonski, Marcin Michalak, Henryk Josinski, and Konrad Wojciechowski. Detection of tumor tissue based on the multispectral imaging. *In Computer Vision and Graphics, Lecture Notes in Computer Science, Part 2, of Springer*, 6375:325–333, 2010.

[32] A. C. Telea. *Data Visualization Principles and Practice*. CRC Press, 2015.

[33] C. Ware. *information visualization perception for design*. Elsevier Inc, 2004.

[34] Chris M. Clark Yong Fan, Nematollah Batmanghelich and Christos Davatzikos. Spatial patterns of brain atrophy in mci patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline. *NeuroImage*, 39:1731–1743, 2008.

[35] Xin Zhao and Arie E Kaufman. Multi-dimensional reduction and transfer function design using parallel coordinates. In *Volume Graphics*, 2010.