Some Recurrent Concepts of Quality Evaluation in Medical Visualization

Beatriz Sousa Santos, Jean Louis Dillenseger*, Jean Louis Coatrieux*

* Laboratoire de Traitement du Signal et de l'Image, Un. de Rennes I, France

Resumo **- Apresentam-se neste trabalho alguns conceitos que parecem fundamentais no problema da avaliação da qualidade da visualização em geral e de dados médicos em particular.**

Abstract **- Some concepts that seem fundamental for the quality evaluation of data visualization in general and of medical data in particular are presented.**

I. INTRODUCTION

Scientific Visualization can be seen as the several processes to transform complex data into a graphical visible and understandable form. The goal is to provide new insight through visual methods. This research area is leading to several visualization methodologies. However, the question of the evaluation of quality of a specific representation remains one of the main challenges of this research area. Concerning the problem of quality evaluation in data visualization in general and in medical data visualization in particular, three different concepts appear to be fundamental: the type and level of components used to convey to the user the information contained in the data (representation of information), the type and level at which evaluation can be performed and the methodologies used to perform such evaluation. In the following sections we will briefly address each of these issues in the scope of the evaluation of visualization techniques which are central to any visualization process.

II. A MODEL FOR VISUALIZATION AND INFORMATION REPRESENTATION LEVELS

A visualization technique is responsible for generating and manipulating a graphic representation from a set of data, allowing investigation through user interaction. Visualization in science, engineering and medicine properly encompasses much more than graphic representation, it involves gaining understanding and insight of the problem solving process and an important point to make is the fact that data fed to a visualization technique is typically sampled from some underlying physical phenomenon which is intended to be visualized and understood, not the data itself.

According to Brodlie et al. [1], there are three distinct operations in a visualization technique:

- (1) the construction of an empirical model from the data, that can act as a representation of the physical phenomenon
- (2) the selection of some schematic means of depicting the model (mapping)
- (3) the rendering of the image on a graphics display.

These steps define the basic structure of a visualization technique. The first operation corresponds to the construction of an internal model of the physical entity from the data (for instance, in the case of generating contours from height data, given as a set of scattered points, a continuous function interpolating the data is constructed). This step involves different aspects of mathematics (sampling, interpolation, approximation).

In the second operation, the empirical model is represented as some abstract visualization object[1]. Usually several abstractions can be used. Some paradigms like the Natural Scene one [11] can formalize this operation. This means here associating the model with understandable 3D forms, e.g. a surface. The choice of this abstraction should be made so as to learn most about the underlying phenomenon.

In the last operation, the abstract visualization object is realized as a graphic sequence; the appropriate graphical primitives are generated together with attributes (in the mentioned example, specifying how areas should be rendered). Finally, the view for the scientist/engineer/ doctor is constructed on the graphics display surface.

It must not be forgotten that, in order to help the user to understand his data, several processes are involved in the transformation from data to graphics display, and all these transformation steps may introduce error or artifacts not present in the data. This means that:

- great care is always needed to minimize those errors and/or artifacts
- methods for evaluating error are fundamental in order to achieve the first goal.

We can perhaps conclude that this is, after all, a process similar to many others in engineering where, in order to use models successfully, it is necessary to be able to

[1] The term abstract visualization object is used to discribe an image object in time and space that is the result of the visualization map, before rendering it into an image.

control the introduced error, i.e. control the approximation

The result of each of the three above referred operations, involved in any visualization technique, can possibly be considered as three different levels of representation of the phenomenon underlying the data. The first level being the empirical model obtained from the data, the second the schematic means resulting from the chosen mapping (i.e. the abstract visualization object) and the third being the final image rendered on a graphic display.

III. EVALUATION TYPES

When evaluating any visualization technique, in order to obtain a measure of the quality of that visualization, the fundamental question that seems to need an answer is "how well does the final image represents the underlying phenomenon and helps the scientist/engineer/doctor to understand it?" This question apparently involves two aspects: (A) the evaluation of the representation of the phenomenon by itself (first part of the question) and (B) the evaluation of the performance of the user in his task when using the visualization, which implies the understanding of the phenomenon, (second part of the question). These two aspects correspond to two different evaluation levels ("low level" and "high level").

It seems plausible that evaluation of visualization techniques should be performed at both levels (high and low) for each of the three levels of data representation (1), (2) and (3) shown in section II, which would result in 6 different evaluation types, corresponding to the following three pairs of questions:

how well:

A1**-** does the empirical model approximate the physical phenomenon to be understood?

A2**-** does the abstract visualization object represents the empirical model?

A3**-** is the visualization object realized by graphic primitives that will produce the final image?

and:

B1- the construction of an specific empirical model from the data

B2- the selection of a schematic representation form

B3- the rendered image

helps the user to understand the underlying phenomenon (and perform his task) ?

Finally, six different evaluation types are necessary.

Generally high level evaluation are made possible only through human observers and low level evaluation could be performed with more objectives and quantitative procedures (this seems to be the case for A1 and A3). However this does not seem to be the case of the low level evaluation of the abstract visualization object (A2). In fact, this second level of representation seems to be the most complex to evaluate, which is due perhaps to the fact that it is the one who gives a larger contribute to the higher level of the fundamental evaluation question Possibly**,** the evaluation of this second representation level corresponds to assess what could be called "perceived quality" of a visualization, i.e. some measure of how successful that visualization is in conveying to the user the information contained in the data and thus in helping him to understand the underlying phenomenon and perform his task.

The other two levels (A1 and A3) seem to give a larger contribute to the low level part of the first part of the fundamental question being more closely related to what could be called the "intrinsic quality" of a visualization.

IV. EVALUATION METHODOLOGIES

Evaluating visualization techniques (of data in general and of medical data in particular) confronts researchers with many issues; however two fundamental issues that must be precisely defined in order to have a evaluation methodology are the type of data and the kind of methods to be used in such study. The kind of test data that could be used is briefly described, as well as three types of evaluation methods and a possible taxonomy of these methods, based on some dimensions that seem adequate to their classification. The adequacy of these evaluation methods to each of the evaluation levels, referred in the previous section, is also addressed.

A. Test data

Basically, two fundamental choices can be done: synthetic or real data [2,3]. Ultimately, the evaluation of a visualization technique should be done with real data, but it is reasonable to begin by using fully specified and systematically controlled data structures embedded in synthetic data. The use of computer generated data results is flexible and allows the detection of errors and inaccuracies of the visualization technique to be evaluated in a way much easier than using real data. In some applications it may require a lot of modelization and may be only approximated, however it is the only method that allows a complete knowledge of the "ground truth". Notice that using synthetic data is a way of short-circuiting the error introduced by the first operation in visualization (modeling the underlying phenomenon), since in this situation that phenomenon is completely known. This can be important to perform evaluation at the other representation levels.

B. Evaluation methods

Several alternatives to evaluate the quality of visualizations seem to exist; each being adequate to perform certain types of evaluation (of certain representation and evaluation levels). For reasons that will be presented later, it seems reasonable to consider the possibility of using the following three types of evaluation methods:

- <u>Methods involving human observers:</u> which rate sets of visualizations, allowing the computation of some quality measures, analogously to what is done to image quality evaluation [4,5], or in ROC studies [6].

• Quality indices: obtained directly from some kind of measures that seem relevant to the quality of the visualization, computed directly from the application of the visualization technique to the data [7].

• Digital observers: which could use models of the Human Visual System (HVS), such as the ones described in [8,9] to estimate ratings that human observers would attribute to visualizations.

Methods involving observers are perhaps the only ones that allow to assess, in its full extent, the human performance on the evaluation of visualization techniques or on using them to perform a task, since they do not use any model, but the "real system"; however they are generally very time consuming, expensive and usually do not provide an easy way to see how quality varies with various parameters of the visualization technique.

Thus, as it is the case in medical imaging evaluation [10], it should be very interesting to have a digital observer that could, in some circumstances, substitute for the human observer; its use would have the great advantages of being much faster to apply and less expensive, when compared to the previous methods; however its applicability would be more limited.

Both referred types of methods would include the "filtering effect" of the Human Visual System, which is generally most desirable, since all visualizations are meant to be used by a human observer, nevertheless the development of what was called "quality indices" should also be considered, if not for other reasons, at least for the fact that these indices, along with some Human Visual System model, could be the base for the development of "digital observers", the last of the three proposed types of evaluation methods.

It should be noticed that methods involving human observers can and should be used to evaluate the performance and validate HVS models as well as digital observers and that digital observers should be developed based on some adequate quality indices in conjunction with HVS models. Thus, the three types of proposed evaluation methods along with real and simulated test data seem to establish an interesting working base to a researcher engaged in the difficult task of developing evaluation methodologies for visualization techniques.

C. A taxonomy for the evaluation methods

Quantitative or qualitative, objective or subjective methods can be considered in the referred three types of evaluation methods.

It is also possible to have methods that take into consideration the Human Visual System (HVS) and methods that do not. Based on these "dimensions" which seem adequate to the classification of evaluation methods, it is possible to propose a taxonomy.

With the purpose of clarifying what is each of these dimensions, let us define them:

• Qualitative/ Quantitative - related to the type of result yielded by the evaluation method: quantitative (having a magnitude that can be and is denoted by a numerical expression), versus qualitative (which cannot or is not)

• Subjective/ Objective - related to the way the result is obtained by the evaluation method: subjective (through the judgment of human observers), versus objective (solely from the nature of the data and visualization method, without the intervention of the observer's judgment)

• Filtered / Not Filtered by the Human Visual System - taking into account the response of the HVS or not.

It should be noticed that the second and third dimensions must not be mixed up, they correspond to different characteristics; an evaluation method may not be subjective in spite of taking into consideration the HVS response; for instance, the methods we have called "digital observers" should use models of HVS to compute some results (possibly quantitative albeit it should not be impossible to produce a qualitative result) without the intervention of the judgment of any human observer and thus they are objective evaluation methods.

Another dimension which could be considered in a taxonomy of the proposed methods is related to the type of observer's perceptual/cognitive processes (of different levels and nature) used in the quality evaluation of visualizations. For instance, if an observer is asked to choose from a set of visualizations which one has less noise, joggles or blur, the type of perceptual/cognitive processes used seems to be of a different level and nature than when the user is asked to give an interpretation of the visualization. These experiments could integrate methods meant to evaluate what we could, perhaps, call "visual quality" and "cognitive quality", respectively, these would be evaluations at different levels the former being at a lower level than the latter. According to this nomenclature, visual quality would be concerned only with the quality of the image (in the sense which is used in Digital Image Processing) and cognitive quality would be concerned with the information conveyed to the observer by the visualization. The referred dimension was included in the taxonomy shown in figure 1, as Visual/ Cognitive, for the sake of completeness; however, due to the complexity of the involved processes, it seems currently exceedingly difficult to develop digital observers to compute predictions about "cognitive quality".

The development of the referred methods is expected to involve much research work since no ready to use methods seem to exist currently. Possibly a sensible approach would be to implement first the methods based on human observers and the quality indices, leaving "digital observers", for a later approach, since these seem harder to develop (which is due to the complexity of the Human Visual System and consequently to the complexity of the corresponding models).

D. Evaluation methods and evaluation levels

It doesn't seem possible to use all types of methods in each of the six evaluation types when evaluating a visualization technique.

	Quantit. / Qualit.	Subjectiv. / Object.	HSV filt. / not filt.	Cognitive / Visual
Methods Human Obs.	X / X	X /	X /	X / X
Quality Indices	/ X	/ X	/ X	- / -
Digital Observers	X / X	/ X	X /	? / X

Figure 1- Taxonomy of evaluation methods (- means not applicable)

Evaluation types A1 (corresponding to the question: how well does the empirical model approximates the physical phenomenon?) and A3 (how well is the visualization object realized by graphic primitives that will produce the final image?) seem to be appropriately performed by objective, quantitative, not involving the HVS methods, i.e. methods we would classify as "quality indices" and which would consist in estimating some kind of error.

Evaluation types A2 and B2 (corresponding to the questions how well does the abstract visualization object represents the empirical model and helps the user?) apparently can be performed only through methods involving human observers, rating visualizations as in image quality evaluation and using ROC curves, respectively. Theoretically, it could also be performed using "digital observers", however as already mentioned, the complexity of the involved cognitive processes and corresponding models, seems to preclude the development of such methods.

Finally and systematizing, it can be noticed that evaluation more concerned with what we have called "intrinsic quality" (types A1 and A3) can be performed using quality indices; evaluation more concerned with what we have called "perceived quality" (all other types) can be performed using methods involving observers (either human or digital). Perceived quality has two aspects related with what we have called "visual quality" and "cognitive quality"; the former being adequately evaluated using methods similar to the ones used in image quality evaluation (ratings obtained using panels of human observers) and the later demanding methods (as ROC curves) which evaluate user performance in executing some task. The evaluation of cognitive quality through digital observers seems out of the question, at least for the moment, due to its complexity.

	Human Observers	Quality Indices	Digital Observers
A1		X	
B1	X		X
A2	X		?
B2	X		
A3		X	
B3	X		X

Figure 2 - Evaluation methods applicable to each evaluation type

V. CONCLUSIONS

In this work three concepts that seem to be recurrent whenever thinking in evaluating data visualizations, in science, engineering or medicine, were analyzed and developed. These concepts are: levels of information representation, types of visualization evaluation and evaluation methodologies. Three levels of representation were found; combining those levels with two levels of evaluation, six types of evaluation were defined. To perform all these evaluation types, methodologies are needed. These methodologies involve test data and evaluation methods. Three different types of methods were presented and a taxonomy was proposed. Finally, the adequacy of the evaluation methods to the evaluation types were briefly addressed.

The ideas presented about each of these concepts, besides being far from fully developed, are controversial; their validity must still be assessed.

REFERENCES

[1] Brodlie, K., L. Carpenter, R. Earnshaw, J. Gallop, R. Hubbold, A. Mumford, C. Osland, P. Quarendon, *Scientific Visualization, Techniques and Applications*, Springer Verlag, 1992

[2] Levkowitz, H., R. Pickett, R. Smith, M. Torpey, "An environment and Studies for Exploring Auditory Representations of Multi-dimensional Data", Grinstein, G., H. Levkowitz (eds), *Perceptual Issues in Visualization*, Springer Verlag, 1995 pp. 47-58

[3] Van Clynenbreugel J., Nuyts, R. Verbeek, M. H. Smet, G. Marchal, P. Suetens, "On the Problem of Validating Computer Assisted Medical Image Interpretation", *Contemporary Perspectives in Three-Dimensional Biomedical Imaging*, C. Roux, J.L. Coatrieux (eds.), IOS Press, 1997

[4] Kaukoranta, T., P. Franti, O. Nevalainen, "Empirical Study on Subjective Quality Evaluation of Compressed Images", *SPIE Conf. vol. 2663, Very High Res. and Quality Imaging*, 1996, pp. 88-96

[5] Avadhanam, N., R. Algazi, "Prediction and Measure of High Quality in Still Image Coding", *Pro. of SPIE Conf., vol. 2663, Very High Resolution and Quality Imaging*, 1996, pp. 100-109

[6] Selin Y., *Detection Theory*, Princeton University Press, Princeton, New Jersey, 1965.

[7] Sousa Santos, B., J. Nunes, C. Coimbra, C. Ferreira, A. M.Tomé, N. Lau, "Quantifying the Objective Quality of Voxel Based Data Visualizations Produced by a Ray Caster: a proposal", *Proc. of Inf. Visual., IV'97* , IEEE Computer Society Press, 1997, pp 111-116

[8] Daly, S., "The Visible Differences Predictor: An Algorithm for the Assessment of Image Fidelity", *Digital Images and Human Vision*, A. Watson (ed.), MIT Press, 1993, pp. 179-206

[9] Lubin, J., "The Use of Psychophysical Data and Models in the Analysis of Display System Performance", *Digital Images and Human Vision*, A. Watson (ed.), MIT Press, 1993, pp. 163-178

[10] Barret, H. H., "Objective Evaluation of Image Quality", *2nd International Summer School-Berder Island*, Brittany, France, June 1996

[11] Robertson P. K., "A Methodology for Choosing Data Representation ", *IEEE CG&A*, 11, 3, May 1991, pp. 56-66.