Visual modeling in an analysis of multidimensional data

A A Zakharova, E V Vekhter, A V Shklyar and A J Pak
Tomsk Polytechnic University, avenue of Lenin, 30, 634050, Tomsk, Russia

Abstract. The article proposes an approach to solve visualization problems and the subsequent analysis of multidimensional data. Requirements to the properties of visual models, which were created to solve analysis problems, are described. As a perspective direction for the development of visual analysis tools for multidimensional and voluminous data, there was suggested an active use of factors of subjective perception and dynamic visualization. Practical results of solving the problem of multidimensional data analysis are shown using the example of a visual model of empirical data on the current state of studying processes of obtaining silicon carbide by an electric arc method. There are several results of solving this problem. At first, an idea of possibilities of determining the strategy for the development of the domain, secondly, the reliability of the published data on this subject, and changes in the areas of attention of researchers over time.

1. Tasks of modeling
Computer visualization is a convenient and common way of presenting information. A consequence of this was the desire of researchers to use this method in solving different scientific and practical. As a result of the systematic and active participation of visual objects in procedures for obtaining the results of scientific research, there is a need in transition from an exclusively auxiliary role of visualization to its justified use as an independent scientific tool [1].

Definition of the visual data model as a visually perceptible image correlates to this data according to some rule. As a model, this image must meet a number of requirements [2]. First of all, visual models are a realized representation of the information defining them, they become an example of a sensible perception of knowledge, which should allow them to analyze, store and process.

The artificial nature of the object, obtained as a result of the system of principles of comparison, implies the preservation of the possibility of subsequent transformations or manipulation of properties of the object. Visual transformability determines the functionality of models as a tool for research [3]. For problems of visual analysis, this property is embodied in the form of the possibility of studying the visible representation of the quasi-reality of model space with the involvement of a natural and well-understood perceptual apparatus.

2. The problem of modeling
The necessity of comprehension and the ability of the researcher to manipulate the space of the visual model for the fulfillment of its purpose face the problem of multidimensionality of the studied data. In this case, authors of the visual model should overcome the contradiction between the need to reduce the dimension of the data and to preserve visual interpretability of the model, which is determined by capabilities of users [4]. Possible solutions of this problem in visual analysis are the search for a variant of the decomposition of the original problem using data of lower dimensionality or the choice of a visual representation that creates a higher dimensionality of the model’s space due to the
attraction of additional visual (color, movement) or mental (memory, perception) informative components.

The created visual model always corresponds to the original only in the part of properties. The investigated object can be associated with a set of models, each of which is determined not only by the properties of the original, but also by the task of the research. In addition, the model, being an artificial analogy, is always subjective. This means the existence of an authorial factor that manifests itself in influencing the way of representation a real or imaginary object [5].

The subjectivity of visual modeling and perception are reserves, the use of which can have a positive effect on achieving the intended result of modeling. In overcoming the problem of modeling multidimensional data, this circumstance can be an effective factor, but requires justification and a preliminary study of the problem, the data properties and the target audience.

3. Requirements for a visual solution

Solving the problem of visual analysis of data leads to obtaining new information that can have a reverse effect on the user's opinion, and therefore there are a number of requirements for the procedure for obtaining an answer of the research question. For example, the features of using visual models are associated with the concentration of attention on the object of studies. This leads to the need for simultaneous visualization of both the background information and the intermediate information needed to answer the research question.

Changing the visual model during the analysis of the data creates the need to verify the correspondence of the new shape to the conditions of the initial analysis task. The phase of such assay also becomes a part of the visual image and changes its perception. The output from an infinite cycle of verification of the received solutions is achieved by means of the visual model itself. For this purpose, the following characteristics have been distinguished:

- Predictability of the consequences of decision making. A significant influence on the speed and quality of the solution is provided by the addition of the visual model by means which demonstrate expected results of various solutions. The most important is the form of representation of the solution that fulfills the function of emotional persuasion.
- Compliance with additional criteria. The question of the research can be formulated in such a way that there are a sufficiently large number of answers. Just one would be chosen as the correct solution which meets the requirements coming from external criteria, in relation to the task under study. Probability of such progress of analysis for a specific task should be taken into account while creating a model, because there is a requirement to present external information using the linguistic and expressive means of the existing model.
- The uniqueness or originality of the solution. The condition for obtaining a solution which is characterized as a new one, is unusual, different from others in some property, but is not inferior to all others in terms of effectiveness. This condition is typical for tasks in which the user's opinion becomes the main criterion for verification. In this case, the model should have means of simultaneously presenting information about the set of possible solutions to the user in a not complicate for the analysis process form.
- Dimensionless of the representation. In the visual model there might be several shapes of data or possible variants of solutions obtained. It allows the researcher to form his opinion, relying on the relative rather than the absolute values in the data. This corresponds to the peculiarities of visual perception, for which a direct comparison of the properties of the observed objects is characteristic. In addition, this makes it possible to exclude the attribute values of the data from the researcher's attention, increasing concentration on solving the main problem and reduction the analysis time.

4. Visual model of multidimensional data

Basing on formulated requirements, authors have developed a methodology for constructing and software tools for creating a visual model designed to solve operational analysis problems of a
significant amount of experimental data. In general, this problem is characterized by the following features:

- The objective of the analysis is to identify data of low reliability.
- The data being examined is a set of descriptions of objects, which belong to a given domain. The number of objects can be more than 10, i.e. it is known to be greater than the number of objects simultaneously available for comparative observation.
- The description of each object in the source data is a set of characteristics (multidimensional data). Sets of features of individual objects may differ, i.e. intersect only partially.
- Types of variables that are separated attributes can be either numeric or categorical.
- The time for obtaining the solution to the analysis problem is limited.

The obtained visual model of multidimensional data corresponds to the conditions of the problem and the requirements to the solution:

- Computer implementation of the visual model (Figure 1) has the ability to represent data of different dimensions and types, creating a visual space for their comparative analysis [6]. Data is imported in text format, which simplifies the exchange with third-party applications.
- The desire to shorten the time of visual analysis leads to the appearance in the structure of the model of the means for automatically constructing the visual image using predefined expressive means. This makes the formalization of the analysis process stronger.
- Using the built visual model is regulated by the user interface, which allows changing both the amount of data involved in the analysis and the metaphor of the presentation. The combination of analysis tools in the space of the visual model, reduces the working time in 3–4 times.
- In solving the problem of representing multidimensional data, decomposition of data into individual components selected by the user and dimensionless display of variables take part simultaneously. This allows to use similar images for 5 types of variables, creating comfortable conditions for their simultaneous study.
- The principle of visual grouping, supported by graphic elements, combines individual variables into a single observable object. This solution preserves a holistic perception of the amount of data pertaining to a particular research object. In the example above, the data dimension is bigger than 10.
- In the comparative analysis, conditions to achieve the modeling goal while simultaneously studying a large number of objects were created (134 in a test study, limited only by computational capabilities).
- The study time is determined by the speed of building the model. In a test task the average time is 6 minutes.
- The model allows evaluating data features ("focus points", "excessive deviations") without determining their formal characteristics at first.
- Model management allows making quick cluster analysis, formulating hypotheses describing internal patterns.
5. External requirements to the visual model

There are many external conditions which are not relevant to the essence of the analysis, but it could affect on its results [7]. Such factors are significant for the activity of the researcher, for the use of his cognitive potential or to control the user’s attitude to the task. It can be an essential part of the visual model and has an impact on its functionality. The developed technique of construction of visual models is focused on improving the effectiveness of the analysis under controlled use of the following factors:

- Limitation of analysis time. It is not a mandatory requirement to have a high decision-making speed for any model and therefore it becomes a resource for using the model. In case of urgent obtaining the result, the type of model and the ways of interaction with it can significantly change. Reducing the time of analysis is achieved through the use of simplified presentation metaphors and the ability to exclude non-informative objects from the perceived image.

- Variability of the data. In some issues, a constant change of the state of the initial data causes a change in the type of their study. For example, in addition to the speed of analysis, it may be necessary to take into account not only the current values of the studied parameters, but also the entire history of changes. The ability to represent variable data is formed when you enter time into a set of object’s features and build groups of object’s images.

- Attraction of subjectivity. If the characteristics of the research need to be addressed to the opinion of the user, including personal experience, emotional evaluation or associative thinking, this leads to an additional personification of the model aimed at activating the respective abilities of the user. The use of subjective experience is achieved by including in the initial data objects that correspond to previously achieved successful solutions.

6. Conclusion

The search for a common approach to solve the problem of visual research of multidimensional data allows using visual data models as an effective research tool. The basis of this decision is a methodical approach to constructing a visual model and using the patterns of visual perception. Additional prospects for the use of visual models are associated with the full and valid attraction of such computer modeling capabilities as simulation of movements, formation of abstract images and the high speed of their obtaining.
Involvement of the individual user's thinking, his personal knowledge and experience is necessary at all stages of visual analysis, including for the formation of a hypothesis of the solution and for making a decision about its reliability. The use of resources of subjective perception and dynamic controllability of the state of the visual model can provide an opportunity for an operative study of multidimensional data and the construction of hypotheses, which is necessary for finding solutions to analysis problems.

The project has been completed as a governmental task № 2.1642.2017/ПЧ – a project about “Cognitive methods of visualization and analysis of multidimensional data during modeling of nonlinear dynamic systems”.

References

[1] Chen C 2013 Mapping Scientific Frontiers (London: Springer)
[2] Bondarev A E and Galaktionov V A 2015 Multidimensional data analysis and visualization for time-dependent CFD problems (Programming and Computer Software) 41(5) pp 247–252. DOI: 10.1134/S0361768815050023
[3] Zakharova A A and Shklyar A V 2016 Visual presentation of different types of data by dynamic sign structures Scientific Visualization 8(4) pp 28–37.
[4] Chen C 2005 Top 10 unsolved information visualization problems IEEE Computer Graphics and Applications 25(4) pp 12–16.
[5] Averbukh V 2013 Semiotic Approach to Forming the Theory of Computer Visualization [Semioticheskij podhod k formirovaniju teorii komp’juternoj vizualizacii] Scientific Visualization 5(1) pp 1–25
[6] Zavyalov D A, Zakharova A A, Shklyar A V and Bagutdinov R A 2017 An integrated approach to modeling by an example of a landfill of disposal of liquid oil waste (Software systems and computational methods [Programmnye sistemy i vychislitel’nye metody]) pp 22–30
[7] Eppler M and Burkhard R.A 2007 Visual representations in knowledge management: Framework and cases Journal of Knowledge Management vol 11(4) pp 112–122