Visualization in Data Reconstruction Tasks

A.V. Shklyar[1,A], A.A. Zakharova[2,B], E.V. Vekhter[3,A]

[A] Institute of Control Sciences of Russian Academy of Sciences
[B] V.A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, Moscow, Russia

[1] ORCID: 0000-0003-4442-7420, shklyarav@tpu.ru
[2] ORCID: 0000-0003-4221-7710, zaawmail@gmail.com
[3] ORCID: 0000-0003-0604-0399, vehter@tpu.ru

Abstract

Many application tasks of multidimensional data analysis which describe the state of real physical or other systems face with difficulties. This is a consequence of the low-quality source data, including missing values, the probability of errors or unreliability of measurements. Incomplete data can become an obstacle for research using many modern informational methods. The current work examines the potential and capabilities of visual analytics tools for preliminary preparation, correction or complete analysis of primary data volumes.

A promising area of application of the approach discussed in the study is the targeted use of visualization capabilities as a data analysis tool. The implementation of specialized visual metaphors is used to solve problems of processing and interpreting data, the sources of which are cyberphysical systems of different complexity levels. Such systems operate in an autonomous or partially controlled mode. A characteristic feature of these systems is the presence of a large number of sensors that collect various types of data. Such data differ in the capacity of the corresponding information channels, their speed and reliability. Examples of such cyberphysical systems are unmanned aerial vehicles (UAVs), robotic stations, and multimodal monitoring systems. These systems can function in conditions where it is difficult to obtain objective observation experience (deep-sea robots). The effective use of data collected by cyberphysical monitoring systems is a condition for solving a large number of application and research tasks.

Keywords: visual model, data reconstruction, metaphor, data model, interpretation.

1. Setting the Task of Visual Research

Many existing approaches to studying and analyzing data of various origins and complexities have a critical dependence on the quality of this data (completeness, reliability, errors) [1], [2]. In the absence of complete data volume, for example, the dynamic system being investigated by the user, these issues can create difficulties and necessitate adjustments to the research methodology. The adjustments may be aimed at clarifying research task or at compensating for existing challenges due to the use of resources that were not previously employed in the analysis process.

Examples of expanding data analytics capabilities through an instrumental approach include a variety of visual research techniques. The purpose of these tools is the increase of analytical process effectiveness through the efficient combination of computational, information, and cognitive resources available to researchers [3], [4], [5]. The means of research in these techniques are visual data models that may differ both in variants of the employed visualization metaphors and methods of communication between users and initial data. This is implemented using interface elements of visual model. Thereby, the visual data model function as an interactive high-tech tool for solving data analysis problems, both as an autonomous tool and as a component of information system [6], [7], [8].

The partial lack of values in the initial data may be the result of different reasons. In general, any values may be attributed to the missing data if they raise doubts or contradict other parameters. In this case, research task is divided into two stages: a preliminary examination of initial data and subsequent analysis. At the second stage, issues of assessing the significance of missing data and reconstructing the necessary information elements are raised. An attitude is also formed towards the results of further analysis techniques applied to studying the initial data volume and data supplemented with the results of recovery process. The objects of this work are visualization tools that allow to find answers to questions from both stages, or, at least, to conduct an expert assessment of the initial data.

During the visual model formation for incomplete data, it is possible to rely on the following heuristics: the value of one object within the analyzed sample of parameter falls within the value range and characterize this parameter, but it belongs to other objects. In other words, the uncertainty of missing values is limited (to some precision) by values present in the initial data.

This assumption imposes significant limitations on the range of issues that can determine the visual data analysis task. The purpose of analysis cannot be the search for abnormal values, extreme points, errors, etc. The most justified direction for efforts of visualization tool developers should be considered the design of tools that allow to perform analytics and form a holistic view of studied data. This representation is understood as a unified object of perception – a visual model. Its interpretation guides the analyst to the reasons for appearance of specific values in the data under study. For example, a visual model of multidimensional data, based on the idea of parallel coordinates, for representing initial data in a single space describes the current state of studied system, with parameters containing descriptions of over 300 informative elements. The representation metaphor makes it easy to identify objects with similar properties, while the degree of "proximity" and its criteria are determined only visually (Fig.1).
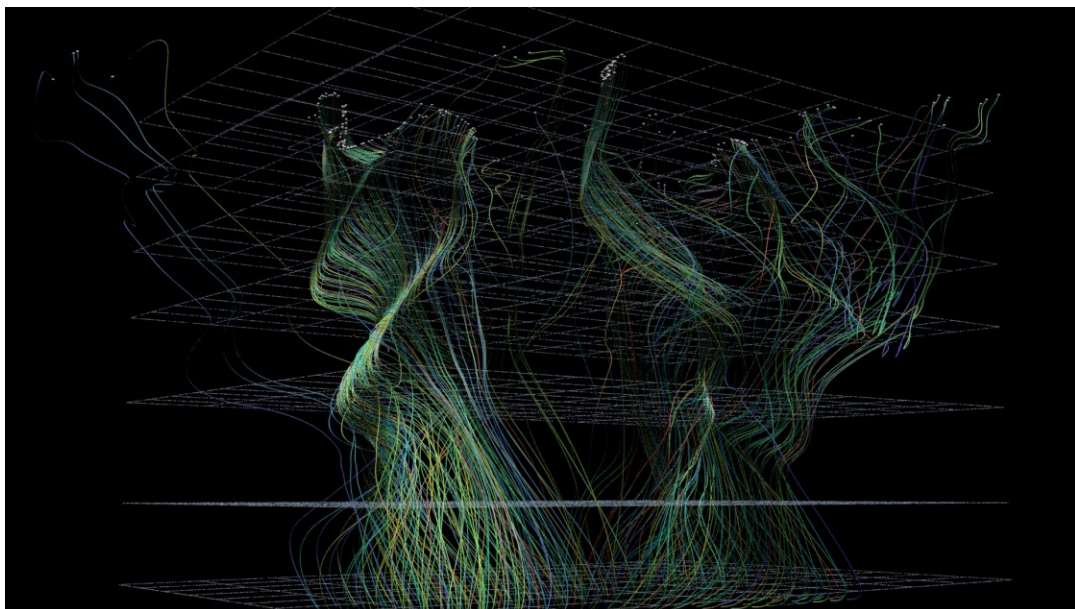

Fig. 1 A visual model of multidimensional data using the idea of parallel coordinates

Some assumptions about properties of such an image, which acts as a model object and is intended for studying and conducting cognitive research, are acceptable [9], [10]. Sets of experimental data characterizing objects of studied field may differ in terms of their volume, acquisition conditions, and states of the object systems or their individual properties, reliability, etc. During preliminary research, the analyst's task may be in obtaining general value judgments that do not contradict either the available data or those that may appear later. Any redundant information in this situation complicates researcher's work. Therefore, the re-

quirement for a visual model is to simplify perception of the overall data picture without losing its informational content.

Within the framework of the analytical solvable task, determining elements of source data or their representation that have excessive information content is an independent issue. The answer to this question becomes an essential part of interpreting initial data. In the developed methodology of visual research, this stage is mandatory and its passage is ensured by the availability of interactive features for visual analysis tools that are necessary for user to make a decision about excluding redundant data from the field of view. In case the negative interpretation results are achieved, the data will be returned to work [11], [12], [13].

2. The Metaphor of Differences

In case of forming only a general idea of the studied system, i.e. without examining secondary data, elements of initial data may be excluded from perception: event identifiers, variable values, or repeating scales (Fig.2). Identifiers are of interest at the stage of comparing data sources and the features of their changes. However, for the initial analysis this information may be relevant if the events on which attention needs to be focused are known. An independent simplification technique is replacing smooth segments connecting data points in the description of each informative element with linear visual connections. In this case, the visibility of presentation is enhanced, but analysis of the entire volume of initial data is hindered by a large number of visible and distinguishable informative images (data points, angles in the images of elements that are concentrators of attention).

The values of individual parameters may also be redundant information, since in most cases the basis for obtaining an evaluative judgment is the dynamics of changes (or lack thereof) of these. It can be assessed by comparing descriptions of events [14], [15]. Consequently, the joint analysis of several event descriptions can be more effective than the sequential research due to better concentration of the user's attention. An experimental evaluation of the user's decision-making speed on the significance of individual information objects when using the parallel coordinates metaphor confirms the possibility of applying visual models with a simplified visualization metaphor (Fig. 2). The time required to complete the expert assessment stage is reduced by 5-15%.
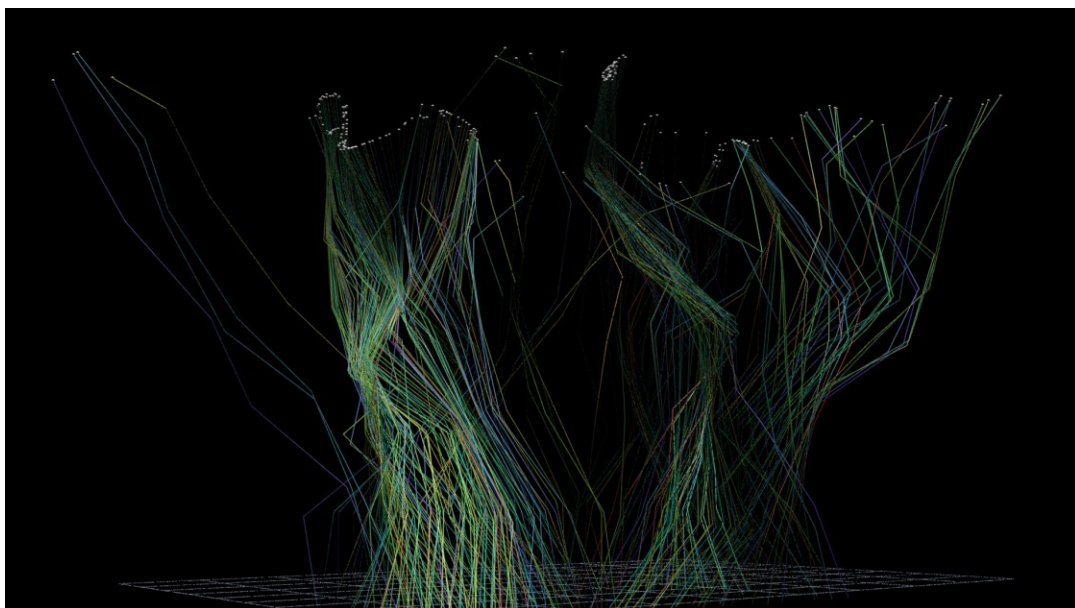

Fig. 2 Variant to simplify the visualization metaphor

Visual representation of multidimensional data inevitably generates a variety of cognitive distortions. They are associated with the simultaneous observation of data images in a shared visual space. Their scaling does not correspond to the values of their own scales or units of

measurement, but rather to the peculiarities of the representation area [16], [17]. A promising solution could be the transition to free user scaling of visual images of individual dimensions. This would allow to exclude (temporarily or permanently) some or all measurements in multidimensional source data from the visual field of the scale, as well as change their scales according to the researcher's current needs.

3. Facilities of the Difference Model

Optimization of the researcher's field of view is associated with the exclusion of secondary data and partially limits the variety of issues available for discussion. In particular, the analysis of values present in the source data becomes independent of the absolute values of the parameters and is based on relative values. A special case of this statement of the purpose for the data examination stage may be the requirement for the researcher to form a holistic view of the properties of the data being analyzed [18]. For example, an evaluation of data quality as a ratio in the initial data volume of complete and incorrect object descriptions (Fig. 3). In this example, a version of the metaphor for visualizing multidimensional data is developed for a preliminary assessment of the quality of investigated data. Each line corresponds to the image of a multidimensional information object, represented in the "spherical" version of parallel coordinates. Data without gaps forms a line that is completely located on the surface of the visual model "sphere" (colored lines, each representing one information object). Information objects with partially missing values have fewer projection points on the spherical surface of the model, and therefore easily stand out in the general data image (blue lines). This situation can be effectively used if there is the possibility to automatically scale the visual representation. However, it destroys the understanding of the range of observed value changes. The importance of maintaining this understanding is not a requirement for visualization tools, since the magnitude of the range of changes is typically known or can be communicated to the researcher using minimal visual methods.

Comparative analysis based on matching of visually observed features in the source data reduces the total analysis time (up to 5%, depending on the data itself). A useful result of visualization in this case is the prediction of parameter values that are missing in the initial data. The unavailability of individual values or significant data fragments may be the result of losses, measurement errors, or researcher's area of interest being beyond the capabilities of available data sources. In each of these situations, predicting missing values is based on the peculiarities of the researcher's visual perception. Implementation is carried out by means of visual analytics, and, importantly, based on the user's subjective knowledge associated with experience in solving similar analysis tasks.

The data reconstruction performed in this way is similar to the formulation of hypotheses explaining the origin of the entire data volume. This can be either an independent task or an intermediate stage of a broader study. It is important that the persuasiveness of visualization, which is one of its most significant advantages, does not become an obstacle to critical verification of the formulated hypotheses. A convenient, although not always possible, solution to this issue may be a comparative analysis of not the initial data itself, but the consequences of adopting different hypotheses visualized as alternative options in the data model [11].
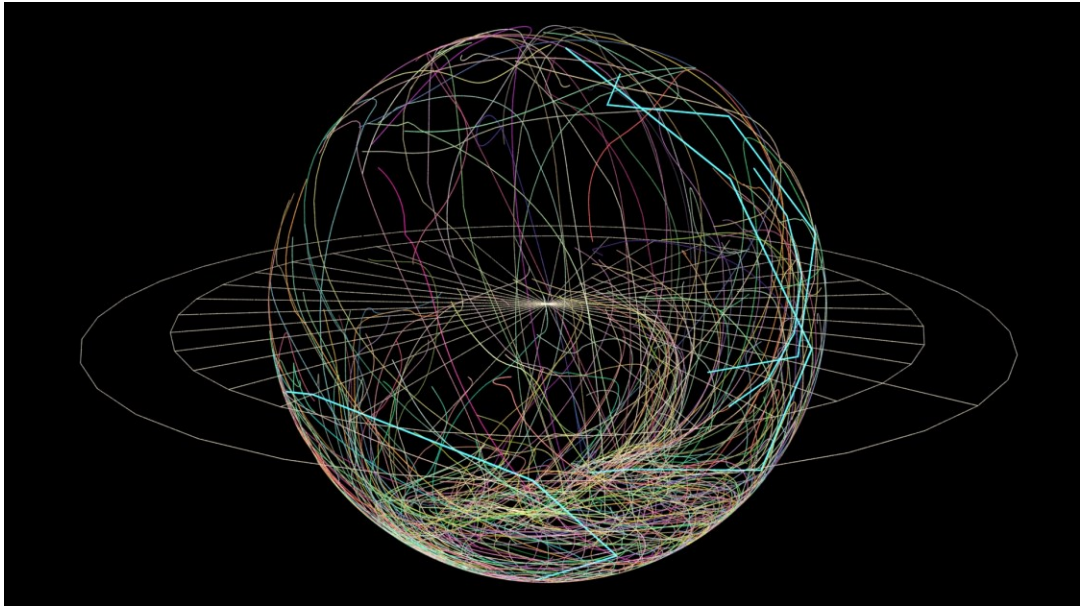
Fig. 3 A metaphor of multidimensional data visualization for preliminary assessment of the data quality

4. The Metaphor of Deviations

The basis for evaluation and necessary reduction the conflict of visualization techniques used in the general perception space may be the allocation of independent basic visual attributes with minimal mutual influence. In other words, it is necessary to understand the compatibility of informative flows in the visual perception process, which can be painlessly separated at the stage of their interpretation. The study of applied visualization systems, including tools used in business analytics and scientific visualization, allows to notice that there are very few known and used ways to display mutually compatible data [19]. It is quite easy to identify two large groups of visualization techniques that can be used independently and jointly: interpretation of the dynamics of changes and deviations from conditional equilibrium.

Techniques based on the observation of dynamic of changes in the data under study (Fig. 4) are characterized by a high concentration of observer attention and the rate of hypothesis generation. In experimental versions of the visual model, a dynamic metaphor was used to preliminary assess initial data characteristics. In this, representing the image of each informative object is a sequential animated process of the image appearance. Each point corresponds to the state of one informational object present in initial data. Point position within the visual model space is defined as the resulting vector in the spherical analogue of parallel coordinates (on the right). The alternative variant (on the right) is distinguished by a dynamic demonstration of the sequence of object states, which significantly exceeds the static version in information content and rate of preliminary analysis.
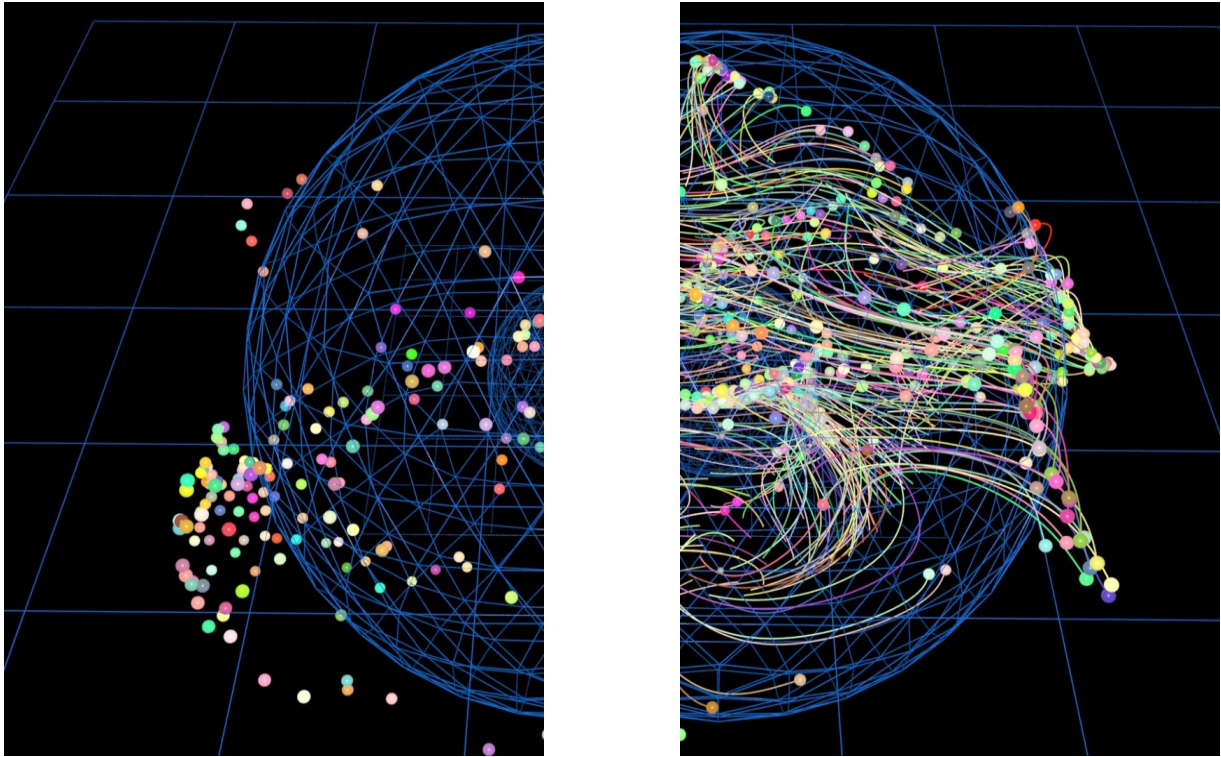
Fig. 4 Experimental versions of the visual model for a preliminary assessment of the initial data characteristics

The general image of initial data, which becomes more complicated over time, is well interpreted if there is a small number of objects (less than 50), for data with a low dimension (the number of parameters in one description is up to 10). In general, interpreting the volume of initial data requires special user training.

Besides, the observation of a changing data image, in which the dynamic component may be not only time, but also any other variable initiates and supports the user's mental activity associated with the reconstruction of missing or otherwise unavailable data. Experimental evaluation of this approach's advantages is complicated by the need to define and systematize ideas about a potential user. In some measurements, when observing dynamic models, preliminary study time was reduced by 9-12% compared to static visual models.

A feature of the reconstruction process for this type of data is the equal employment by the user to generate a new hypothesis of both the results of the current interpretation and data previously available to the researcher. This feature specifies the field of application for visual analytics tools and imposes requirements on the user's knowledge and experience. This means that the amount of information a person can freely manipulate and work with a dynamic model has natural limitations. Developers of visual representation metaphors should take this into account.

The idea of interpreting deviations has significant number of implementation versions. One of the most popular is the assessment of spatial deviations in data images from a given initial state. Other versions of the same idea are deviations in the color representation from a predefined neutral state, as well as any other difference between the instance of visualized data and what is expected. This variety of versions determines the popularity of these visualization tools. However, this very often leads to obtaining visualization metaphors that do not take into account the patterns of visual perception. Examples of such erroneous metaphors from the group with deviations may be common versions of data color coding, when the peculiarities of color and brightness perception are not considered, or the visual disproportion of scales for all or some parameters in the multidimensional data array. This determines the presence and appearance of visual accents.

5. Advantages of the Deviation Metaphor

The advantage of the deviation group representatives can be considered in obtaining a more accurate interpretation of data presentation in the following cases: a significant number of analyzed events persists for large data volumes; in the presence of chaos in the values of studied parameters; difficult understanding without the use of additional visualization techniques. This can be explained by the lack of restrictions in the duration of observing the visual model properties, as well as the possibility of simultaneously displaying a variety of data in the visual field. The interpretation procedure for such an image has significant differences from the models of dynamic metaphor group, as it becomes possible to adapt the speed of cognitive processes to individual actual user needs (Fig. 5). Version of the deviation metaphor is proposed to assess the advantages of this type of visual models. This version is designed to form a general idea of the initial data, as well as to decide on the possibility of borrowing missing values from information objects with similar properties (yellow and green lines). In the proposed model, differences in information object's properties are visualized as a deviation accumulated with distance from the model center. Therefore, the detection of similar objects does not cause difficulties.

Potential practical applicability of visual models, using the metaphor of deviations, is higher than that of data-oriented dynamic or static models. This is caused by the effective use of natural visual matching mechanisms that are familiar to most users and have a high speed. According to experimental estimates, reducing the time spent by the observer on allocating objects of increased interest based on differences in their properties from the total initial data volume or explicitly specified "reference" values reaches 25-40%. It also depends on the ability to match properties of a particular representation metaphor and scale of studied deviations.

Highlighting these two groups of data visualization techniques does not make it mandatory to choose one of them when developing visual analytics tools. There are many applied visualization systems that combine both approaches. For example, dashboards in BI-systems combine different display options. They are focused on emerging changes (KPI monitors). Techniques for representing dynamic changes in the form of static images have also become popular (for example, the widespread metaphor of "Japanese candlesticks" - Candlestick Chart [20]). Presentation of information in the form of dynamically changing KPIs is oriented on analyzing not the values themselves, but the moments and magnitudes of their changes. Since simultaneous monitoring the value of several variables creates a significant burden for the user [21], digital values are often supplemented with graphical elements to reduce it. They indicate the direction of recent changes, but it also complicates the overall image of data. The metaphor of "Japanese candlesticks" is conceived as a compact and rich form of data representation that is constantly changing. Therefore, it has become a popular tool for solving forecasting tasks. The disadvantage of this approach should be considered additional requirements for training of specialists using these tools, as well as the limited amount of information transmitted to the observer at one time. This is a consequence of transition to an unusual sign system.
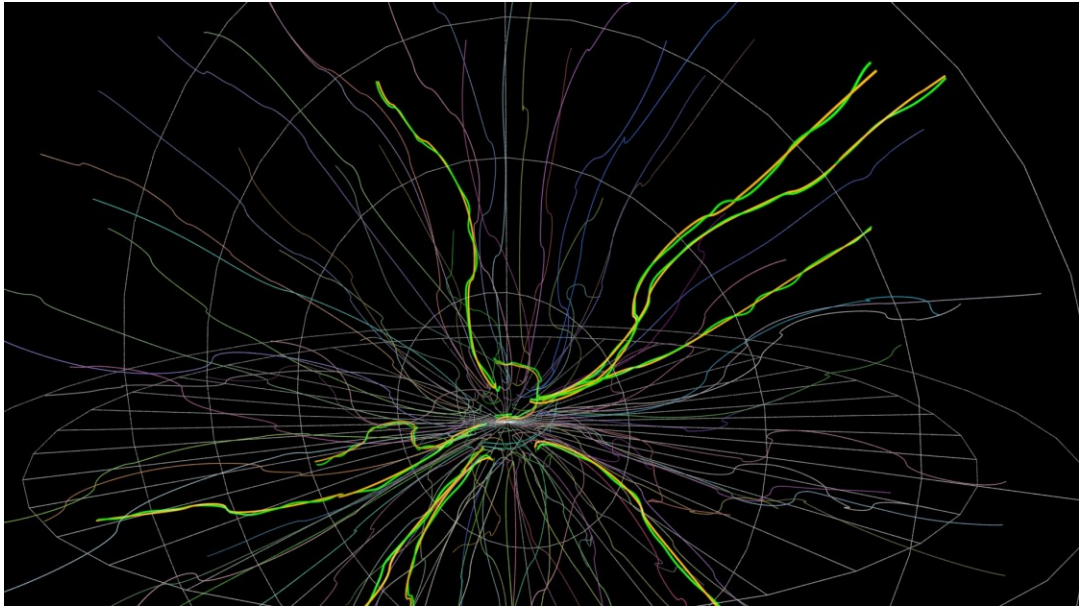
Fig. 5 Variant of a visual model based on deviation assessment

6. The role of perception patterns

It worth to note that for representatives of each specified group of visualization techniques, as well as for hybrid variants, the statement regarding target use of visual perception patterns is true [22]. Understanding the role of specific perception features in searching for answers to various questions will allow development of visual analytics tools based on understanding and planning principles of their operation. Unfortunately, there are many examples of incorrect attitudes towards the application of visualization tools in practice. This raises questions about the value of visualization in general.

For the tasks of reconstructing missing data, the applicability of visual perception patterns has its own characteristics, and ignoring them reduces the effectiveness of visual analytics tools. For example, the principle of perception integrity manifests itself in the form of an observer's unconscious desire to mentally combine individual visual elements into a group. This occurs when there are assumptions about the presence of a general rule defining their appearance. In the reconstruction task, this creates conditions for formation hypotheses about similarity between the values of individual parameters. It happens if general descriptions of information objects in the visual model field are similar to each other. In the absence of other grounds for recovering missing data, this principle may provide basis for further analysis of the system under study.

The integrity of perception can also manifest itself through assumptions about the existence of information structures. These structures combine similar elements in the visual field and have their own properties that also need to be studied. However, this is not always fair or necessary (Fig. 6). For example, a visualization metaphor is useful, if it allows to simultaneously represent both the studied data and areas of acceptable values. This metaphor is formed as a result of interactive model management and can lead to the perception and interpretation of a visual structure that is not intended for this purpose. A convenient solution to this challenge is decomposition of the visual model. This will allow studying of its individual components. The introduction of such control changes is aimed at obtaining additional information by the user. It is necessary to focus on specific values (features of the initial data, the detection of which, using traditional data pre-processing techniques, is possible if a research hypothesis is provided). The proposed metaphor is a tool for forming statistical hypotheses tested at the next stage of data analysis.
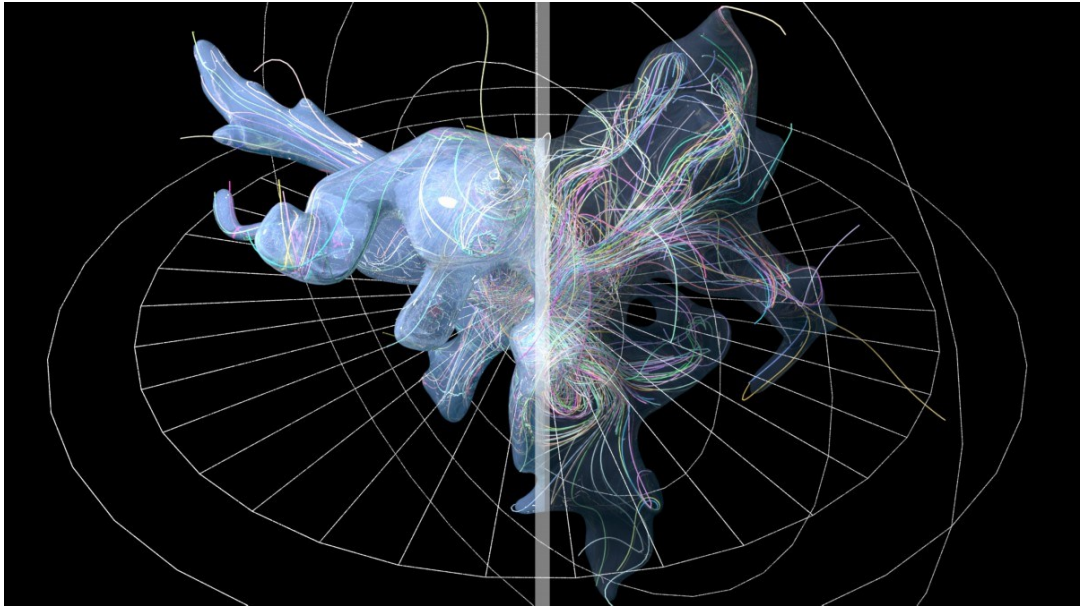
Fig. 6 A visual model of multidimensional data that allows simultaneous presentation of the volume of data under study (animated tracks - images of information objects in the space of the visual model) and variations in visualization when making any permissible changes to the data image (blue translucent areas)

An equally significant feature of visual perception is connected with the principle of generalization. It largely determines the processes of recognizing objects in visualized information or its individual fragments. In accordance with this principle, recognition of known to the observer object occurs even in the presence of distortions in transmitted data or other variants of information noise. In other words, recognition of an object ends with determination of its belonging to a specific class of objects that belongs to the user's knowledge system. The result of generalized perception is the ability to operate with information about not individual objects, but their corresponding classes. This allows restoring gaps in the initial data when features of objects are combined at the classification level. In the metaphor using the idea of parallel coordinates (Fig. 1), a reduction in research time by 5-10% is explained precisely by perception generalization.

The principle of objectivity manifests itself in creating prerequisites for dividing visual information flow into following data: relevant to the observer's current interest; elements that belong to the environmental group according to the criteria established by the user. In visual analysis tasks this principle corresponds to decomposing the visual field into objects. Their parameters and values are perceived as independent information structures. Existence of such structures can be explained in the course of ongoing research. In addition to the objects, there are signs of relationships between the selected structures and informative environment that influences the state of objects of interest. Objectivity of visual perception is the basis for persuasiveness of observed images and, to some extent, the data used to create the image. Therefore, the process of data analysis is initiated, since objectivity is a confirmation of the cognizability and reproducibility of studied events.

In addition, the opposite side of the objectivity principle is the possibility to detect objects – informative elements of the visual model that have unusual or atypical properties (Fig. 3). This complements results of interpreting the observed visual model state by raising new questions and corresponds to managing the process of analyzing initial data (including its completion or changing the goal). In discussing data recovery tasks in descriptions of individual information system elements, the objectivity of perception can play a significant role. Since this is a factor in convincing the user about the reasonableness of initial data and possibility of their extrapolation to descriptions of elements that lack necessary values. The absence of such conviction causes repeated cycles of verifying intermediate interpretation results. Their num-

ber exceeds what is necessary and depends on the subjective readiness of the researcher to accept controversial and non-obvious hypotheses.

The constancy of visual perception becomes a pattern that is of great practical importance for visualization tasks in general, as well as issues of using visual analytics tools for data recovery and, in particular, data analysis. Visual representation of multidimensional data, based on the metaphor of complex geometric images in three-dimensional space, explicitly uses the perceptual constancy to provide the user with ability to interactively control the model space. It also allows the user to control the sequential change of observation points for data image or productive refinement of metaphor. The principle of constancy ensures the user maintains understanding of changing data model image. It is frequently used to organize the possibility of an arbitrary direction for analyzing multidimensional descriptions of objects. Moreover, it takes into account the subjective distribution of interest areas in the data space.

In the process of designing visualization tools, the principle of perception constancy imposes restrictions on the range of allowable changes in the data image. Transformation of the perceived data image, any systematic conversions of the representation metaphor are independent techniques of cognitive research. They are aimed at providing users with ability to search for interpretation results that correspond to questions posed. However, according to the principle of constancy, exceeding the limits of permitted transformations results in changing the class to which an object of perception (or its individual parts) was assigned. This may be a desirable outcome if changes in the visual model were made to form a qualitatively new interpretation hypothesis. It can also be a negative phenomenon if the change in class of informative object has caused destruction of the system of facts, connections between them and the corresponding knowledge previously acquired by user.

Considering and applying the principles of visual perception mentioned above is an obligatory aspect of designing visual analytics systems. This allows researchers to make actions and conclusions systematized. It does not mean that the cognitive process will lack such a useful component as insight. However, its achievement becomes dependent on the accumulation by the users of their own analytical activities results. In addition, an important stage in the cognitive process is its completion. It refers to the state when the answer to the initial research question has been obtained, and there is sufficient confidence in the correctness of this answer to move on to a new question and a new research goal of the degree. The time it takes for a user to reach this state determines the effectiveness of tools and corresponding analysis techniques. This is determined not only by the properties of problem being solved and visual analytics tools, but also by the interaction of these tools with the user.

Discussion of the specific aspects of interaction between the visualized information and perception and researcher's mindset cannot be limited to discussing the applicability of classical perception patterns in the interpretation of visual data models. This is associated with a significant increase in the variety of expressive techniques that can be used in visualization tasks, as well as the emergence of specific visualization tools that are related exclusively to digital technologies and do not have close analogues in reality [23].

7. Interpretation of the Movement

An example of this trend in visualization is the use of animation. Animation is the movement of elements, that constitute a data image composition, or any other metaphor transformations, or changes in the work of methods included in a general metaphor, that are synchronized in the user's perception with the time of representation observation (Fig. 4). Evaluation the prospects of involving animation in visual analytics tasks indicates that judgments, related to the action of perception patterns, are largely applicable to it. However, there are several specific manifestations of the animation's influence on the interpretation of visualized information, which can either enhance or seriously reduce the activity of cognitive process. The problem of animation is associated with active involvement of the subjective movement

understanding. At the same time, there are contradictory properties that become enhanced if visual analytics tool is developed for collective use.

One of the significant patterns, that are used by creators of various type animated images, is readiness of perception for the next movement or expectation, formed as a result of understanding events, that have already occurred. In cinematography, this fact is used to reduce the time required for moving on to the next scene, but in visual analytics, readiness of perception is an indicator corresponding to the viewer's understanding of causal relationships affecting origin of studied data. This understanding may be linked to the user's previous knowledge or new hypothesis, which was verified by the visual model.

Direct consequence of the active role of informed waiting in interpreting a visual data model is the rapid detection of inconsistencies between the expected event development in visual field and the actual behavior of the studied system (Fig. 4). Disturbed waiting is an effective way to detect errors in two directions: in the initial data (it is possible to evaluate the quality of studied data) or in the user knowledge. Additionally, the waiting formed for user in the interpreted movement is an indicator of achieving a certain level of understanding (confirmation of the hypothesis about data origin). Efficiency, in this case, is explained by the fact that there is no need for additional efforts on the part of visualization tool developers to draw attention to these errors, since switching attention to error analysis takes place naturally for the user.

Additional advantage of animation is its active role in maintaining a high level of observer involvement in the study. This refers to the influence of user's psycho-emotional state on the process of solving a visual analysis problem. In other words, prolonged concentration of user's attention at work leads to slowing down or stopping the cognitive process. Animation, as part of the visualization metaphor, can regulate user involvement through such techniques as rhythm of perception, switching between observation and interpretation modes, and drawing personal interest [24].

8. Methodological Approach to the Visual Research Task

In general, methodological approach to the problem of incomplete data research consists of choosing one of possible solution algorithms, preparing conditions for its implementation, and developing tools. They should fit the task and determine means of visual analytics and qualitative evaluation of the result. The following alternative options can be considered as two possible algorithms for conducting research on incomplete data:

A. Elimination of data from the research process that are not available for study in traditional or other ways, due to the lack of values in the part of samples. Creating a data model based on the assumption about insignificant role of missing data in the process of answering questions facing the user. In this case, it is important to use visualization to form the researcher's confidence in the ability to continue solving the problem of analyzing incomplete data sets. The lack of confidence in the correctness of actions can be compensated at the stage of evaluating the obtained analysis results. However, the transition to verifying results is preceded by the hypothesis generation stage. During this stage, the presence of user's subjective uncertainty leads to excessive multiple proposing of new hypotheses. This significantly slows down the solution of analysis task. It should be noted that the potential applicability of this solution algorithm is largely based on the objectivity and generality of visual perception.

B. Replacing missing values in descriptions of informative data elements with assumed ones, i.e. hypothetically possible values. A specific case of this solution is the introduction of a special point - "no data available" on appropriate scales. Next, analysis is carried out, taking into account the ability of state of data sources when they do not generate necessary values. The basis for reconstruction in this variant of the solution algorithm is the hypothesis that individual values of multidimensional data are close to each other. At the same time, the condition of sufficient value similarity in most other available measurements must be fulfilled. This hypothesis is not obvious and can be justified by solving a specific task by visually evaluating

the similarity between complete sets of values for available informative elements (if the requirement of sufficient quantity in the original data set is met). In this case, visualization also becomes a convenient instrument for justifying the applicability of chosen reconstruction approach with quick preliminary analysis. This happens thanks to the involvement of visual perception patterns such as integrity and constancy.

Development of visualization tools, based on these assumptions, is primarily aimed at defining the effective metaphor of visual representation. Characteristic property in this case is the ability to simultaneously display the full volume of available data in the visual field. This is essential for forming a holistic view of the research subject and its features for observer. At the same time, developed metaphors act in similar way in both algorithms that are used for solving analysis problem, i.e., they assess the remoteness (visible difference) or proximity (similarity between images) of informative elements in a multidimensional data space, but have opposite goals.

A. Exclusion algorithm requires a representation metaphor that emphasizes the assessment of diversity. It means that in the case where the diversity of incomplete data has negligible influence on the meaning, it is possible to make a decision about the admissibility of their exclusion from the total volume of initial data.

B. For the replacement algorithm, a visualization metaphor should facilitate an effective search for objects with similar properties in order to provide a basis for borrowing missing data.

It should be noted that in some cases the same visualization metaphor may be applicable for both options, because the requirements for analysis tools are similar. In this case, the correctness and effectiveness of their application will depend on the user's preparedness, their understanding of the visual analysis purpose and subjective effectiveness of the expressive methods used.

At the stage of qualitative assessment of the choice correctness and success of using visualization tools for solving any data analysis task, it becomes necessary to determine the success of overcoming many objective problems and mistakes made by users or developers. Common problems for a wide variety of data and tasks are always the following:

A. Representation of non-numeric data. The traditional method of using categorical measurement scales may cause certain difficulties in situations where several different scales of this type are required to visualize the information element of initial data. Technically, there are no difficulties and the formal procedure for creating an image does not face obstacles. However, for users who want to visually determine the degree of similarity or difference between data elements, this question can become quite difficult, because understanding distance in a semantic space defined by categorical scales rarely corresponds to the user's subjective experience. In this case, the integrity of informative element image formed in observer's perception can become the basis for attributing it to a certain class of objects, and using categorical scales can be replaced by assigning perceived attributes to objects.

B. Assessment of the significance. The ratio in the visual representation of the significance of individual properties is both a reserve of any representation metaphor and its weak side. In this case, it is difficult to determine in advance the factors that control user's attention. In practice, this corresponds to a high probability that significant elements of the data image will escape observer's attention. Alternatively, minor fluctuations in values can cause a deep cognitive process and distort the general meaning of studied data.

C. Abstractness or the metaphor complexity. Geometric properties of a visual image represent a model picture of the source information, and are very rarely directly related to the real properties of the data. The representation metaphor for any multidimensional data, in the perception of an ordinary observer, does not correlate in any way with the perception of data itself. A simple illustrative example of this state can be a regular geographical map, where information about the height of surface point is presented in the form of color coding. In more complex situations, when the studied data has more complex organization or properties what are unknown to the user, any metaphor generates an image with high visual com-

plexity and degree of abstraction. In general, this fact is not problematic, but taking into account perception patterns, unexpected results may occur when the observer determines the category of an object or its informative properties.

Direct consequence of these difficulties is the increasing significance of the experience, perception and cognitive skills of a particular researcher. There is an obligatory stage of determining the features of information communication performed using visual models. The selection of data representation method can be made manageable and controllable in order to delegate users the opportunity for interactive subjective adjustment of the model to their needs or perception.

9. Application of Visual Metaphors for Processing and Interpreting Data from Cyberphysical Systems

A promising area of application for the approach discussed in this work is solving data processing and interpretation problems, that arise from cyberphysical systems (CPS) of various complexity levels, operating in an autonomous or partially controlled mode. These systems have a large number of sensors responsible for collecting many types of data, differing both in the capacity of corresponding information channels as well as their speed and reliability. Examples of such cyber-physical systems/CPSs are unmanned aerial vehicles (UAVs), robotic stations operating under conditions that prevent the acquisition of objective observation experience (deep-sea robots), and many other multimodal monitoring systems.

All these tasks are characterized by the need to obtain analysis results under the following conditions: the impossibility of taking repeated measurements, the presence of high levels of noise, the loss of original data fragments. During evaluating advantages of the approach considered in this work to solving the visual research problems, a set of visual metaphors was developed. It is designed for assessing the quality and preliminary analysis of heterogeneous data. The source of this data is monitoring systems. The test data had different volume and level of complexity: from 20 to 300 records, each of them could contain 5-18 parameters of different types, including measurements of temperature, pressure, gas content, etc. (Fig. 7). Using unified metaphors in relation to the analysis of data obtained by the CPS during environmental monitoring allowed to reduce the time for developing visual research tools. For example, when initial data volume increases (up to 500 records), the time required for creating a visual model does not change. In addition, such unification has reduced the cost of training a visual analyst to work on tasks in new directions.
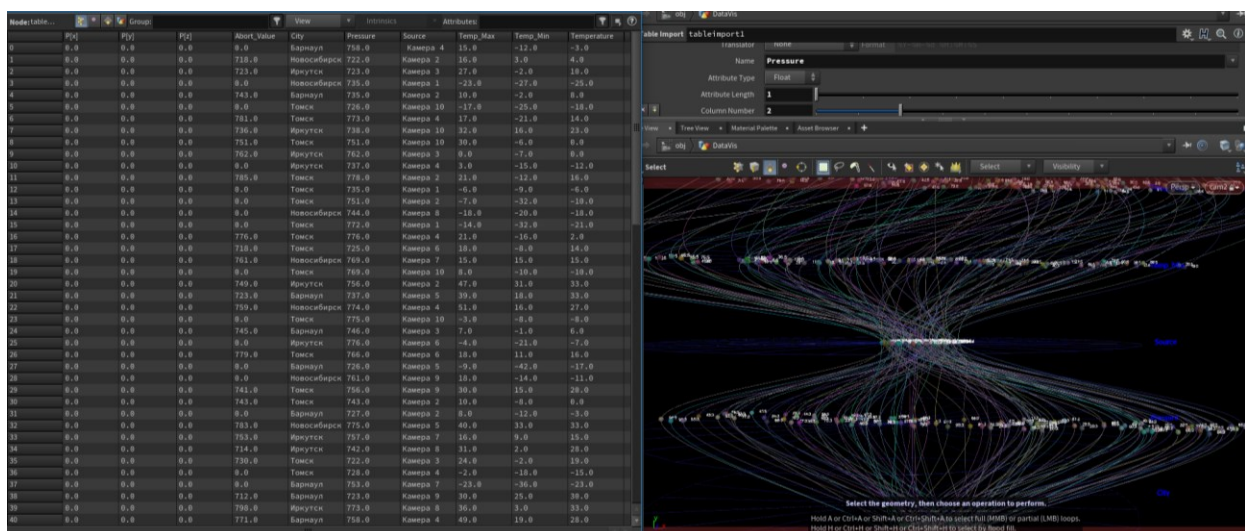

Fig. 7 An example of test data and a visual metaphor chosen for analysis

Usage of the parallel coordinate metaphor for the task in question is caused by the need to draw on the experience obtained by users in solving similar problems earlier. In addition, the

effectiveness of the proposed visualization tool is based on the use of a deviation metaphor, which has shown good results in analyzing multidimensional data that are similar in parameters to the test data. As a visualization technique necessary to indicate missing values, excluding the corresponding model nodes from the user's field of view and changing the brightness of the image fragments of informative elements with missing values are used (Fig. 8). At the same time, the simultaneous presence of a node name and shaded element image fragment in the field of view is used as an active pointer to the model area that requires the user's attention.
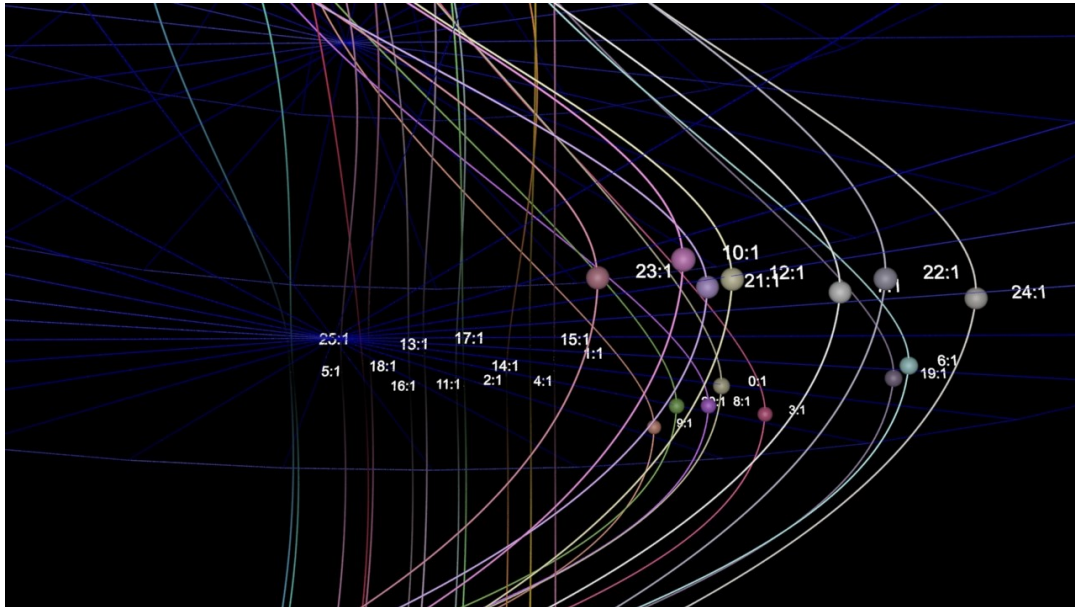

Fig. 8 Visualization of missing values. The name of each model node contains the number of the informative element, as well as the value index in the record

In some situations, for example, when the source data has high noise or a significant number of missing values (this is typical for the analysis of data collected by the CPS), the color-coding method becomes insufficient. This happens because the user's attention becomes scattered during the process of interpreting a heterogeneous saturated image. To reduce the burden on the user's perception, a technique was used that reduces the visual model detailing while maintaining its informativeness (Fig. 9). In the proposed solution, nodes of the visual model with no values are shifted to a special image area (here, the central point). For the user, the pointer to the area of focus no longer on nodes, but rather on intersections between images of complete and damaged elements.

For solving the problem of missing value recovery (when in a series of experiments, there is no complete identity of the set of CPS measuring instruments or technical and other failures occur during recording data), when using the considered metaphor (Fig.10), it is necessary to select images of complete informative elements, with most parameters close in values to images of elements with missing data. To do this, it is enough to select the damaged element and determine number of the informative element suitable for constructing a borrowing hypothesis. In order to simplify the process of creating a borrowing hypothesis, the parameter with corrupted data is shifted to the bottom part of model. This allows to separate complete and damaged elements while maintaining a general understanding of all values available for analysis. The proposed method is particularly useful in situations where the source of direct borrowing is not in the initial data and in order to form hypothesis, a full understanding of the initial data and their features is necessary. This situation is the most common for tasks of data processing collected by the CPS.

Fig. 9 Technique for reducing the detail of a model

Fig. 10 General view of the test data model

Conclusion

In this work, an analysis of numerous factors that have a significant impact on the effectiveness of using visual analytics tools for solving applied problems associated with analyzing multidimensional data of different types was performed. As a test example of such task, the problem of preparing and researching multidimensional data collected by cyberphysical systems of different types was examined. One of the features of such data is the lack of some values in the descriptions of informative objects. A methodological approach that relies on the advantages of visual perception in generating hypotheses of interpretation and subsequent analysis has been proposed.

The possibility of designing visualization tools for solving applied problems has been shown. This includes the interactive determination of the most promising algorithm for solving the problem of researching data of a specified type, as well as for other applied tasks where achieving the research goal is complicated by the lack of proven methods for utilizing the potential of modern information technologies. The role of visualization and patterns of

visual perception is shown in both the process of researching initial data and controlling the effectiveness of using the subjective potential of the user of visual analytics tools.

In the task of reconstructing the data collected by CPS, the metaphor of deviations can be applied at the stage of evaluating data quality when using visual analytics tools repeatedly. This is a typical situation for the analysis of data received from monitoring systems, including those with a large number of different sensors. Validation of the developed research methodology has demonstrated that the user's understanding of expected values, formed in this case, makes the deviation model a tool that speeds up the solution of analysis problem (by 25%, according to experimental estimates). The metaphor of differences provides good results in evaluating the possibility of borrowing values from similar information objects and is therefore also recommended as a basis for designing visualization tools at the stage of preliminary research. In addition, versions of visual models that are based on generalization of perception and interpretation of movement can increase the speed of solving research problems by up to 40%.

References

1. V. Shklyar, A. A. Zakharova, и E. V. Vekhter, «Methods of solving problems of data analysis using analytical visual models», Sci. Vis., V. 9, Is. 4, P. 78–88, 2017, doi: 10.26583/sv.9.4.08.

2. Chen, «Science Mapping: A Systematic Review of the Literature», J. Data Inf. Sci., V. 2, Is. 2, P. 1–40, 2017, doi: 10.1515/jdis-2017-0006.

3. V. Shklyar, A. A. Zakharova, E. V. Vekhter, и A. J. Pak, «Visual modeling in an analysis of multidimensional data», J. Phys. Conf. Ser., V. 5, Is. 1, P. 125–128, 2018, doi: 10.1088/1742-6596/944/1/012127.

4. J. F. Rodrigues Jr., L. A. S. Romani, A. J. M. Traina, и C. Traina Jr., «Combining Visual Analytics and Content Based Data Retrieval Technology for Efficient Data Analysis», 4th International Conference Information Visualisation, IEEE, 2010, P. 61–67. doi: 10.1109/IV.2010.101.

5. A.E. Bondarev, V.A. Galaktionov. Generalized Computational Experiment and Visual Analysis of Multidimensional Data (2019). Scientific Visualization 11.4: 102 - 114, DOI: 10.26583/sv.11.4.09.

6. Vieira, P. Parsons, и V. Byrd, «Visual learning analytics of educational data: A systematic literature review and research agenda», Comput. Educ., 2018, doi: 10.1016/j.compedu.2018.03.018.

7. H. Chen и др., «Uncertainty-Aware Multidimensional Ensemble Data Visualization and Exploration», IEEE Trans. Vis. Comput. Graph., V. 21, Is. 9, P. 1072–1086, 2015, doi: 10.1109/TVCG.2015.2410278.

8. V. L. Averbukh и др., «Virtual reality as an instrument of computer visualization», в 2019 International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON), 2019, P. 0786–0792. doi: 10.1109/SIBIRCON48586.2019.8957854.

9. M. A. Yalçin, N. Elmqvist, и B. B. Bederson, «Cognitive Stages in Visual Data Exploration», в Proceedings of the Beyond Time and Errors on Novel Evaluation Methods for Visualization - BELIV '16, 2016. doi: 10.1145/2993901.2993902.

10. V. N. Kasyanov, Methods and tools for information visualization on the basis of attributed hierarchical graphs with ports, Siberian Aerospace Journal. 2023, Vol. 24, No. 1, P. 8–17. Doi: 10.31772/2712-8970-2023-24-1-8-17.

11. Batch и N. Elmqvist, «The Interactive Visualization Gap in Initial Exploratory Data Analysis», IEEE Trans. Vis. Comput. Graph., V. 24, Is. 1, P. 278–287, 2018, doi: 10.1109/TVCG.2017.2743990.

12. H. Guo, S. R. Gomez, C. Ziemkiewicz, и D. H. Laidlaw, «A Case Study Using Visualization Interaction Logs and Insight Metrics to Understand How Analysts Arrive at Insights», IEEE Trans. Vis. Comput. Graph., V. 22, Is. 1, P. 51–60, 2016, doi: 10.1109/TVCG.2015.2467613.

13. R. Pienta и др., «VIGOR: Interactive Visual Exploration of Graph Query Results», IEEE Trans. Vis. Comput. Graph., т. 24, вып. 1, сс. 215–225, янв. 2018, doi: 10.1109/TVCG.2017.2744898.

14. R. J. Crouser, L. Franklin, A. Endert, и K. Cook, «Toward Theoretical Techniques for Measuring the Use of Human Effort in Visual Analytic Systems», IEEE Trans. Vis. Comput. Graph., 2017, doi: 10.1109/TVCG.2016.2598460.

15. R. Akhmadeeva, Y. A. Zagorulko, и D. I. Mouromtsev, «Ontology-Based Information Extraction for Populating the Intelligent Scientific Internet Resources», в Knowledge Engineering and Semantic Web, A.-C. Ngonga Ngomo и P. Křemen, Ред., Cham: Springer International Publishing, 2016, P. 119–128.

16. V. Shklyar, A. A. Zakharova, E. V. Vekhter, и D. A. Zavyalov, «Visual detection of internal patterns in the empirical data», в Communications in Computer and Information Science, Volgograd, Russia: Springer Verlag, 2017, P. 215–230. doi: 10.1007/978-3-319-65551-2_16.

17. Chen и M. Song, «Visualizing a field of research: A methodology of systematic scientometric reviews», PloS One, V. 14, Is. 10, c. e0223994, 2019.

18. Ware, Foundation for a Science of Data Visualization. 2012. doi: 10.1016/B978-0-12-381464-7.00001-6.

19. Meirelles, «Diagramming as A Strategy For Solving Graphic Design Problems», Просмотрено: 29 февраль 2020 г. [Онлайн]. Доступно на: https://www.academia.edu/2749924/Diagramming_As_A_Strategy_For_Solving_Graphic_Design_Problems

20. S. Nison, Japanese Candlestick Charting Techniques. 2024.

21. J. H. Larkin и H. A. Simon, «Why a Diagram is (Sometimes) Worth Ten Thousand Words», Cogn. Sci., V. 11, Is. 1, P. 65–100, 1987, doi: 10.1111/j.1551-6708.1987.tb00863.x.

22. Branchini, U. Savardi, и I. Bianchi, «Productive thinking: Tlie role of perception and perceiving opposition», Gestalt Theory, V. 37, Is. 1, P. 7–24, 2015

23. Branchini, U. Savardi, и I. Bianchi, «Productive thinking: Tlie role of perception and perceiving opposition», Gestalt Theory, V. 37, Is. 1, P. 7–24, 2015

24. J. Choi, S. Jung, D. G. Park, J. Choo, и N. Elmqvist, «Visualizing for the Non-Visual: Enabling the Visually Impaired to Use Visualization», Comput. Graph. Forum, V. 38, Is. 3, P. 249–260, 2019, doi: 10.1111/cgf.13686.