

Visualization Metaphors in the Tasks of Exploratory Analysis of Heterogeneous Data

R.A. Isaev^{1,A,B}, A.G. Podvesovskii^{2,A,B}, A.A. Zakharova^{3,B}

^A Bryansk State Technical University, Bryansk, Russia

^B V.A. Trapeznikov Institute of Control Sciences of RAS, Moscow, Russia

¹ ORCID: 0000-0003-3263-4051, ruslan-isaev-32@yandex.ru

² ORCID: 0000-0002-1118-3266, apodv@tu-bryansk.ru

³ ORCID: 0000-0003-4221-7710, zaawmail@gmail.com

Abstract

The subject of the study is the construction and application of visual models using the concept of visualization metaphors in the context of exploratory analysis of heterogeneous data. This study considers improved variants of the previously proposed visualization metaphors that can be used as a basis for building visual models. A technology for exploratory analysis of heterogeneous data based on the joint use of different visualization metaphors is proposed. The process of visual data exploration at the stage of exploratory analysis using the proposed technology is demonstrated to be iterative and multiscenary, contingent upon the analysis goals. The software tool developed to implement the proposed technology is described, along with its additional functionality to calculate and export quantitative characteristics of the visual model. The software tool is then considered in the context of exploratory analysis of a synthetic data set. The future direction of the proposed approach to the construction of visual models, the technology of exploratory data analysis and the software tool for its support are determined.

Keywords: exploratory analysis, visualization, visualization metaphor, visual analysis, heterogeneous data.

1. Introduction

Exploratory data analysis is a preliminary analysis that identifies the general properties of data, their internal relationships, patterns, and anomalies. The basic ideas of exploratory analysis were outlined in the classic book [1]. However, some later publications (e.g., [2]) attempted to discuss its provisions and put forward alternative ideas. The results of exploratory analysis usually form the basis for in-depth data analysis.

Numerous applied publications demonstrate the relevance of exploratory analysis in all areas involving the processing of weakly formalized data [3-5].

Heterogeneous data is one type of such data that often requires initial exploratory analysis. This data can be obtained from complex, distributed, heterogeneous dynamic systems, such as cyber-physical systems.

Exploratory analysis of heterogeneous data often involves solving highly abstract and weakly formalized tasks. Therefore, organizing effective exploratory data analysis requires leveraging the cognitive potential of the analyst. The most effective approach to engaging the human analyst's cognitive functions is through visualization and visual analysis capabilities [6]. Numerous studies have demonstrated the successful application of visualization in solving tasks related to understanding objects or processes of various natures. There are known examples of application of visual analytics in problems of computational fluid and gas mechanics [7-9], in solving problems of optimization of parameters of distributed experiments

in high-energy and nuclear physics [10], in software design [11], as well as for text data analysis [12].

To describe and solve the problem of visualizing heterogeneous data, an approach based on the concept of visualization metaphor can be utilized [13]. A metaphor is a set of principles that describes how the characteristics of the object under study, such as a set of data, are transferred into a visual model space, which can be either two-dimensional or three-dimensional. A visualization metaphor consists of two components that are applied sequentially:

- the spatial metaphor determines the type and dimensionality of the visualization space, as well as the arrangement of model elements within it.
- the representation metaphor specifies the characteristics of the visual image, which are necessary for visualizing certain properties of the object under study that are most significant at the current stage of analysis.

Usually, several representation metaphors correspond to one spatial metaphor. The complexity of the object being studied and the need for consistent visualization of its properties and characteristics during research are the reasons for this [14].

Visual data models should be simple and convenient for analysts to interact with, and the concept of cognitive clarity is often used to describe this aspect [15]. This concept refers to the ease of intuitively understanding and interpreting a given amount of data represented in a visual model. Insufficient cognitive clarity of the model often leads to difficulties in comprehending the data, incomplete or erroneous interpretation of certain data elements, and so on. Simultaneously, a high level of cognitive clarity in the visual model during exploratory analysis enables the researcher to quickly identify important properties of the dataset, detect incompleteness and anomalies, and accelerate the identification and interpretation of patterns and internal relationships.

2. Data under study: features and tasks of their analysis

This paper assumes that the data being studied has a general structure consisting of a set of *objects*, each described by a number of *properties*.

The description of each object is an enumeration of the *values* of some or all of its properties.

It is implied that objects belong to the same class or at least related classes. In the first case, this means that all objects have the same sets of properties. In the second case, it means that the sets of properties for different objects may not coincide, but their intersection is non-empty. This situation is one of the possible manifestations of data heterogeneity.

The properties of objects can be measured on different scales, including qualitative (such as nominal or ordinal) and quantitative (such as interval or absolute).

Data heterogeneity can lead to potential incompleteness, which is characterized by missing values for some properties of all or some objects. There are several types of incompleteness, including:

- some objects in the data set may lack a property value due to either insufficient measurement or a complete lack of measurement reliability.
- additionally, due to the heterogeneity of the data set, certain objects may not possess one or more properties that other objects in the same data set have.

Considering the type of data incompleteness is crucial for effective visual analysis. It impacts the validity of conclusions regarding the sufficiency of available data for analysis and the possibility of filling data gaps based on the analysis results.

The following are general classes of research problems that may arise when working with such data.

- Tasks related to the generation of hypotheses about patterns and relationships in the data. For example, a hypothesis might involve whether and how two or more properties are related.

- Tasks related to selecting a subset of data elements (objects or properties) from a common data set based on a certain criterion (or set of criteria). Examples are finding objects with abnormal property values, selecting objects with “good enough” property values, finding the most informative properties, selecting clusters of similar objects, etc.

Because of the potential incompleteness of the data, an additional task that arises in these classes of problems is to assess the sufficiency of the available data to achieve the objectives of the study. Another related task is to fill in the gaps in the available data set with synthetic values that are deemed most plausible by the analyst.

3. Data visualization metaphors for exploratory analysis

The proposed visual analysis technology in this paper can be based on various visualization metaphors and combinations thereof. The main idea behind sharing visualization metaphors is to combine their advantages. This is achieved by placing the visual images of the data in the same visual field, creating a common visual model of the data that can be perceived as a whole by the analyst. Visualizing data simultaneously in different aspects increases its cognitive clarity.

This technology is based on two well-known visualization metaphors, which we will describe next.

3.1. The metaphor of spatial relationships

The proposed main metaphor for visualization in visual analysis technology is the three-dimensional metaphor, as described in [16].

The choice of the above visualization metaphor is conditioned by the fact that it has the following properties useful in the context of the task of visualization of heterogeneous data.

- It implies placing a set of heterogeneous data in a single “dimensionless” visual space. This allows the analyst to simultaneously perceive visual images of initially disparate quantities.

- It organizes the placement of visual images in such a way that the analyst can examine the data from two complementary perspectives: first, as whole images of objects with all their properties, and second, as whole images of properties on the entire set of objects.

- This metaphor remains workable in the case of incomplete data under investigation and, moreover, can be very useful in just such situations.

Modifications to the metaphor are suggested to enhance its ability to visualize different data characteristics. The components of this metaphor, including the spatial metaphor and possible representation metaphors, are described below.

According to the spatial metaphor, the visual model employs a cylindrical coordinate system to describe its space. The base of the model consists of parallel planes, with each plane corresponding to a different data property (Fig. 1). By default, the properties are arranged in a bottom-up direction, with the first property at the bottom of the visual model and so on. If any properties are removed from the display, the visual model will be rearranged to eliminate gaps and remain centered in the visual space.

The concentric circles reflect the various levels of the scale used to measure the characteristic associated with the corresponding property. Fig. 1 visualizes three levels for all properties: the “beginning of the scale” (smallest circle), the “middle of the scale” (middle circle), and the “end of the scale” (largest circle).

It is assumed that all quantitative properties are measured on an interval or ratio scale, so their initial values can be normalized to a dimensionless range from 0 to 1.

Each object present in the dataset under study is assigned a specific angular coordinate in a cylindrical coordinate system. This coordinate ensures uniform placement of objects in the visual model space.

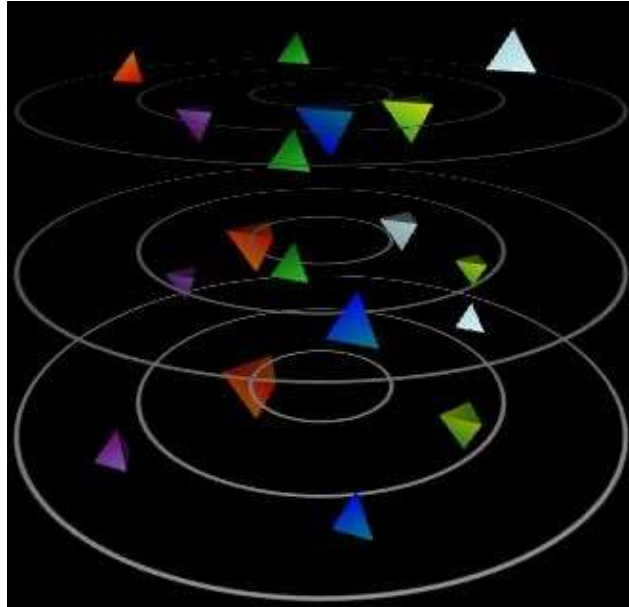


Fig. 1. Basis of visual model and example of visualization of source data

This visualization metaphor supports a number of representation metaphors, some of which can be used in combination.

The basic representation metaphor is responsible for visualizing the basic properties of the underlying data. It uses the idea of visual markers. A visual marker is a graphical object that is placed on one of the planes and associated with a specific data element. This object has the following parameters that can correspond to different characteristics of the data.

- *The position of the marker on the scale.* In the simplest case, it can reflect the very value of a particular property for a particular object. Such a variant of the metaphor is presented in Fig. 1. Another possible variant implies that the position of the marker reflects the deviation of the value of the property of a given object from some value. Such a value can be, for example, the average value of the property for all objects or some reference value of the property.
- *Shape.* In the proposed variant of the metaphor (Fig. 1), the marker has the shape of a tetrahedron. It carries information about the deviation of the property value of a given object from a given value (in the example - from the average value of the property for all objects). The marker direction "up" means superiority over the reference value, "down" - vice versa. Another option is to use the form to visualize the value of some discrete property of an object (it can be interpreted, for example, as an object class). In this case, objects of different classes will have markers of different shapes (tetrahedron, cube, sphere, etc.).
- *Size.* In the variant shown in Fig. 1, the size of the marker is proportional to the deviation of the object property value from the average value. If objects do not differ in terms of some property (the values of this property are equal for all objects), then the markers of objects on the corresponding plane have the shape of cubes. Also, the size of markers can be responsible for visualization of deviations from other reference values or for visualization of initial property values themselves.
- *Color.* In Fig. 1, the color of the marker serves as an object identifier. Other ways of using color are also allowed. For example, color can be used to visualize the value of a discrete property (i.e., to distinguish objects of different classes), as well as to visualize the sign and magnitude of deviation of the property value from the specified value. In the latter case, the sign of deviation is conveyed by means of hue (red/blue or green/red), and the magnitude of deviation by means of color intensity.

Further, all metaphors are described on the basis of the interpretation of marker parameters that corresponds to the example in Fig. 1. It should be noted that a different interpretation of marker parameters may entail a different interpretation of these metaphors.

The visualization metaphor for object profiles (Fig. 2) gives the analyst the ability to see each object as a whole. Visual images of profiles are broken lines, the shapes of which allow to visually assess the similarity or difference of objects in terms of the balance of values of their properties. This enables the detection of pairs and groups of similar objects. For example, Fig. 2 shows that the green and purple objects have similar property profiles (the analyst's eye is quick to note the proximity of these profiles to mirror symmetry), while the orange object differs significantly from them.

The visualization metaphor for property profiles (Fig. 3) depicts each property as a single characteristic of a set of data. A closed profile visualization aids in estimating the amount of variation of property values across a set of objects, as indicated by the degree to which the shape deviates from a regular polygon. The analyst can also identify objects with anomalous property values. In Fig. 3 it can be seen that in terms of the middle property, the objects are equivalent, but in terms of the other properties, the objects differ, with the blue object having an abnormal value on the top property.

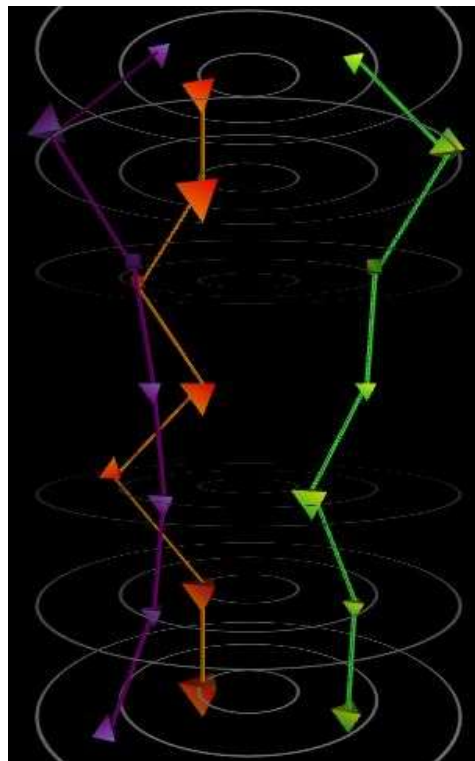


Fig. 2. Visualization metaphor for object profiles

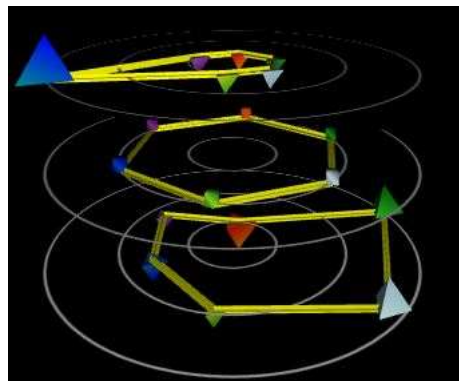


Fig. 3. Visualization metaphor for property profiles

The visualization metaphor for deviation from the reference (Fig. 4) illustrates the extent to which object properties deviate from those of the reference object. The reference object is typically an abstraction that does not exist among the set of real objects. In the example de-

picted in Fig. 4, the reference is linked to the maximum values of all measurement scales (i.e., the reference object has the highest possible values for all properties). It can be seen that the orange object has a high degree of deviation from the reference, while the gray object is quite close to the reference. Other interpretations of the reference object are possible. For instance, one could explicitly state reference values for each property.

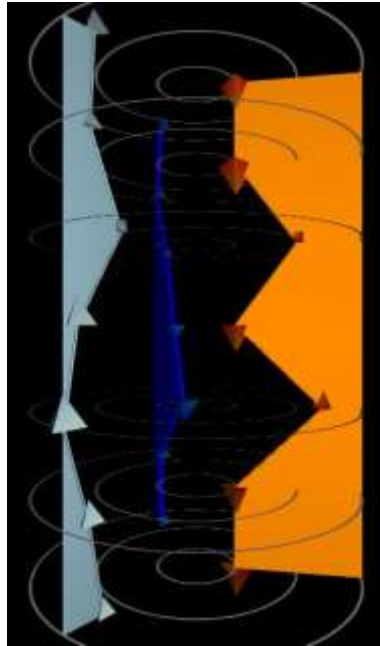


Fig. 4. Visualization metaphor for deviation from the reference

The visualization metaphor for deviation from the mean (Fig. 5) illustrates the extent to which an object's properties differ from those of an “average” object, which is an abstract object with average values for all properties. This can be interpreted as the degree of atypicality or abnormality of the object. The average value for a particular property can be calculated using the arithmetic mean, geometric mean, or median value, depending on the type of measuring scale. In the example in Fig. 5, not only the values but also the signs of deviations are taken into account. It can be seen that the object on the right loses to the “average” object in almost all properties, while the object on the left, on the contrary, surpasses it. A variant of the metaphor that disregards the sign is also feasible.

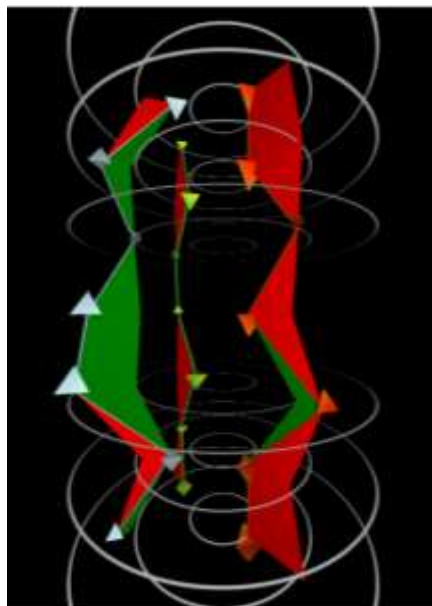


Fig. 5. Visualization metaphor for deviation from the mean

The metaphor for object comparison (Fig. 6) simplifies the visual comparison of two objects by showing how much they differ from each other in each property. In Fig. 6 shows that the object on the left is strongly superior to the object on the right in some properties, while being equal or slightly inferior in other properties. The example considers the signs of differences, but an alternative variant showing only the absolute value of differences is also possible.

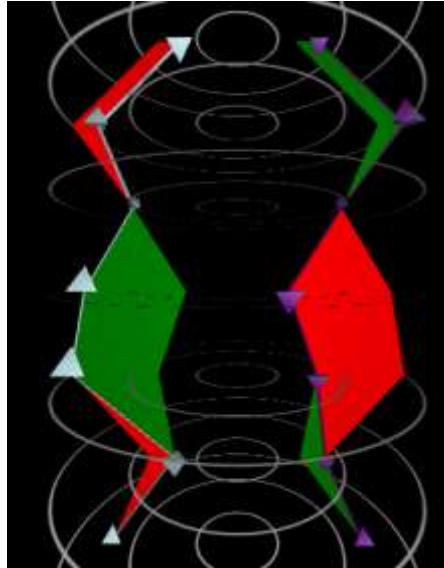


Fig. 6. Visualization metaphor for object comparison

The visualization metaphor for data gaps (Fig. 7) draws the analyst's attention to the presence of value gaps in the analyzed data. The metaphor depicts 'incomplete' object profiles, where sections of profiles corresponding to properties with missing values are visualized by red lines. These sections highlight the gaps in the object description and indicate their location. Such a visual image of the data can be helpful in evaluating the adequacy of the available data for further visual analysis for a specific purpose. In Fig. 7, the purple and green objects have very little known data in their description, and there are only isolated gaps in the description of the remaining objects.

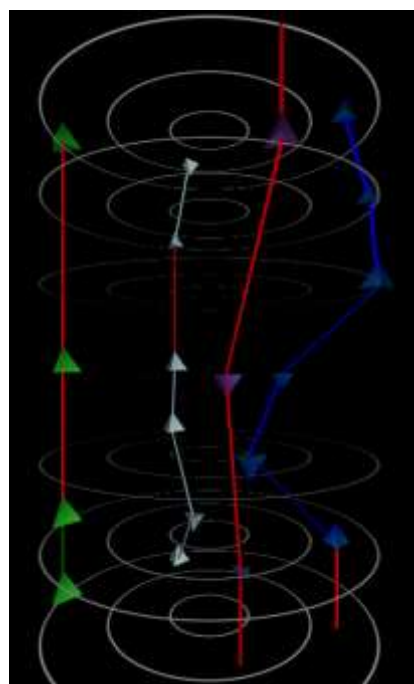


Fig. 7. Visualization metaphor for data gaps

Concluding the description of the spatial relations metaphor, let us note its limitations in terms of the data to be visualized.

- First, the data under study must have the "objects-properties" structure, which is described in the previous section of the paper.
- Second, an obstacle to the use of the metaphor is an excessively high degree of data heterogeneity, which is understood here as a small number of common properties (with a large number of unique properties) among objects. That is, it is a situation in which objects become difficult to compare because they are too different in nature.
- Third, another possible obstacle is excessive incompleteness of data. Although metaphor can visualize gaps in the dataset, the prevalence of gaps over known meanings will render visual analysis ineffective.

3.2. The petal metaphor

The technology can also use a two-dimensional petal visualization metaphor, which was described in [17], in addition to the spatial relationship metaphor.

This metaphor creates distinct visual representations of the objects being studied. The visual representation resembles a pie chart with petals, with the number of petals corresponding to the number of object properties that are being visualized (Fig. 8). Typically, the length of each petal is calculated so that the area of the petal is proportional to the value of the corresponding property, taking into account the normalization of all values to the unit range.

The resulting visual image enables a comprehensive assessment of the object at a glance, including its strengths and weaknesses (if such interpretation of properties is appropriate in the context of the problem being addressed). Also, this metaphor provides an opportunity to compare objects with each other, including the search for similarities and differences between them.

In the above example, color is used to identify an object (images of different colors correspond to different objects). Other interpretations of color are also possible: for example, to identify a property (different properties correspond to petals of different colors), as well as to visualize the deviation of a property value from a given value.

When using this metaphor to visualize incomplete data, it is possible to indicate gaps in the description of an object by means of a special visual sign, as shown in Fig. 9 (information on one of 7 properties of the object is missing). Other ways of indicating incomplete data are also possible, but the general principle should be observed: the display of a gap in the data should not be identical to the display of a property with a null value.

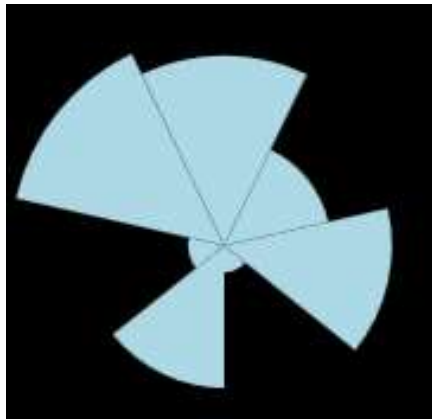


Fig. 8. Example of a visual image based on the petal metaphor



Fig. 9. Representation of gaps in object description based on the petal metaphor

4. The technology for exploratory data analysis using visualization metaphors

We propose a technology for exploratory analysis of heterogeneous data using the considered visualization metaphors, which is presented in Fig. 10. As it follows from the scheme, the application of the technology involves the systematic execution of a number of steps. Each stage has its own level of balance between the role of the analyst and the role of the software in its execution.

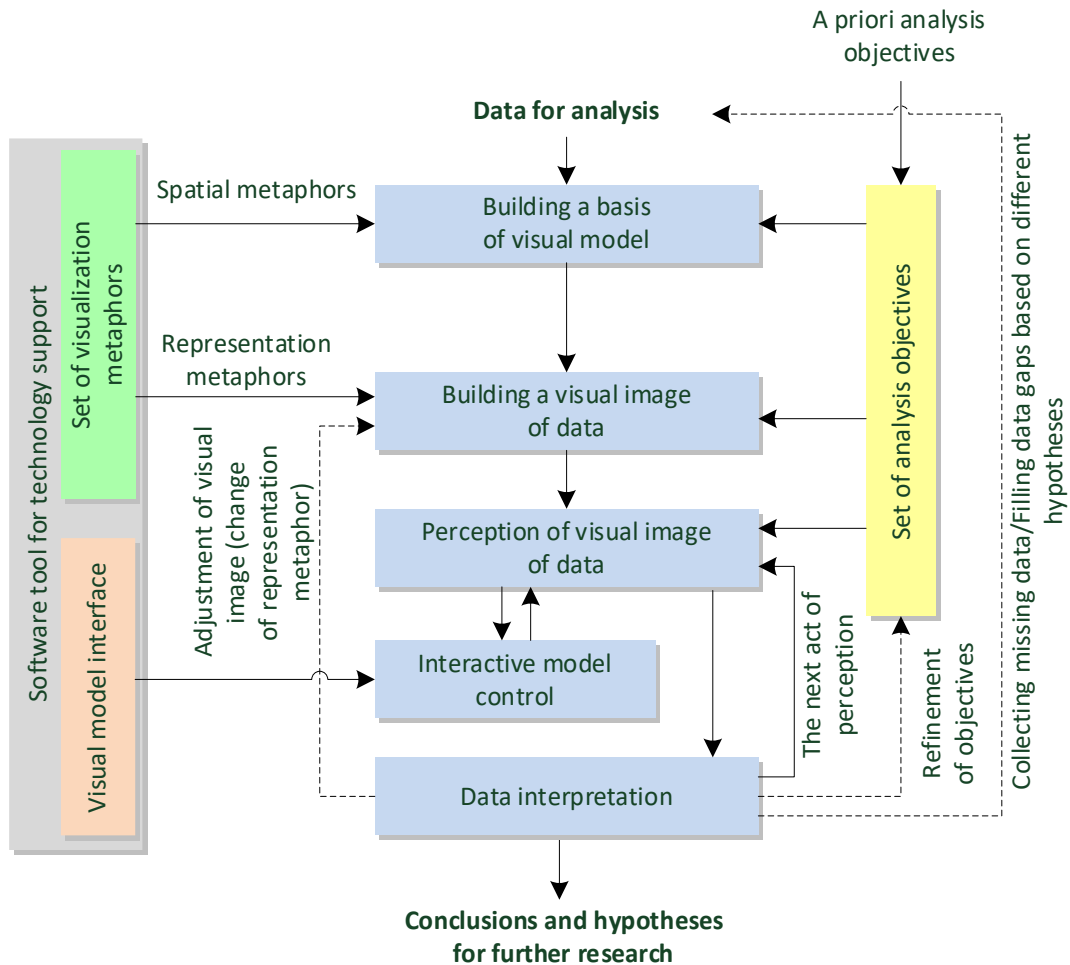


Fig. 10. Scheme of the technology for exploratory data analysis using visualization metaphors

The application of the presented technology is characterized by a significant degree of variability and multiple scenarios. The analysis stages' meaning, the number of repetitions, and the transitions between stages depend on several factors in a particular situation.

- A set of initially set (a priori) analysis objectives. Examples of such objectives are assessment of sufficiency of available data for analysis (this is relevant in case of data gaps), search for objects with anomalous properties, selection of the best object or subset of objects in some respect, etc.

- The results already obtained during the visual analysis process. They can influence, firstly, the necessity and feasibility of the next iterations of the analysis, and secondly, the nature of these iterations (i.e., which variant of the linkage in the technology scheme will be used).

Dotted lines in the scheme indicate not obligatory transitions between stages. Such transitions may never be used during the application of the technology – for example, when solving the simplest analytical problems.

Let us consider the meaning of the cycles in the scheme. They correspond to different variants of iterative actions that are performed in the process of using the technology.

1. The cycle “Perception of visual image of data” – “Interactive model control”. This cycle corresponds to the analyst's routine activities that are associated with solving a specific visual analysis task or subtask. It reflects the analyst's attempts to come closer to understanding the data under investigation by interactively manipulating its visual image. The cycle ends when an understanding (interpretation) of a piece of data relevant to the task at hand has been achieved. The duration of such a cycle (the number of its iterations) may depend on such factors as the quality of software support for interactive control of the visual model, the analyst's preparedness (in particular, the level of mastery of interactive control tools), and the quality of applied visualization metaphors.

2. The cycle “Perception of visual image of data” – “Data interpretation”. This cycle reflects the analyst's solution of a number of similar tasks, which are related to the study of a particular aspect (property) of the data with the help of the chosen representation metaphor. The process of solving one such task is described by the cycle discussed above. The cycle is completed when all tasks within a particular representation metaphor have been solved, i.e. interpretations of the associated data fragments have been achieved. The length of time it takes to complete this cycle naturally depends on the amount of data that is subjected to visual exploration. Because of human cognitive limitations, the analyst will be forced to divide the research process into separate acts of perception.

3. The cycle “Building a visual image of data” – “Data interpretation” – “Adjustment of visual image”. The cycle describes the process of changing the representation metaphors to solve new types of tasks that arise for the analyst in the course of data exploration. The cycle ends when all required types of tasks have been solved (i.e., the goals of the study have been achieved) or when it is concluded that some tasks cannot be solved under current conditions (e.g., due to insufficient data).

4. The cycle “Data interpretation” – “Refinement of objectives” – “Building a visual image of data”. The cycle reflects the possibility of adjusting a priori goals and objectives of data analysis based on intermediate results obtained during the analysis. This is usually accompanied by a change in the representation metaphor due to the transition to a new type of research task.

5. The cycle “Data interpretation” – “Collecting missing data” – “Building a visual image of data”. This cycle describes the situation of interruption of visual research due to the impossibility of achieving all or some of its goals. This may be due to a lack of data, and the analysis can be resumed once the missing data have been obtained.

The application of visual analysis technology leads to conclusions about the studied data set. The form of these conclusions depends on the analysis objectives. For instance, they may describe detected anomalies in the data or a subset of the most preferred objects. One form of conclusions may be hypotheses about the data, such as statistically significant relationships

between various indicators. These hypotheses are subject to further testing by formal methods, such as mathematical statistics.

Another form of conclusion is recommendations aimed at improving the validity of the analysis results. Recommendations may pertain to collecting more data on certain objects or properties.

Additionally, applying technology to an incomplete dataset may result in automated filling of missing values in the original dataset. In this case, values are suggested based on a hypothesis about the distribution of values in the data, formulated by the analyst during visual exploration.

5. Software support for the proposed technology

Software support for the technology of exploratory analysis of heterogeneous data is implemented in the form of a Windows application. This application is developed on the .NET Framework platform using Windows Presentation Foundation (WPF) technology, which allows creating applications with rich graphical user interface.

Fig. 11 shows the software tool interface with a visual model of some data set based on two considered visualization metaphors.

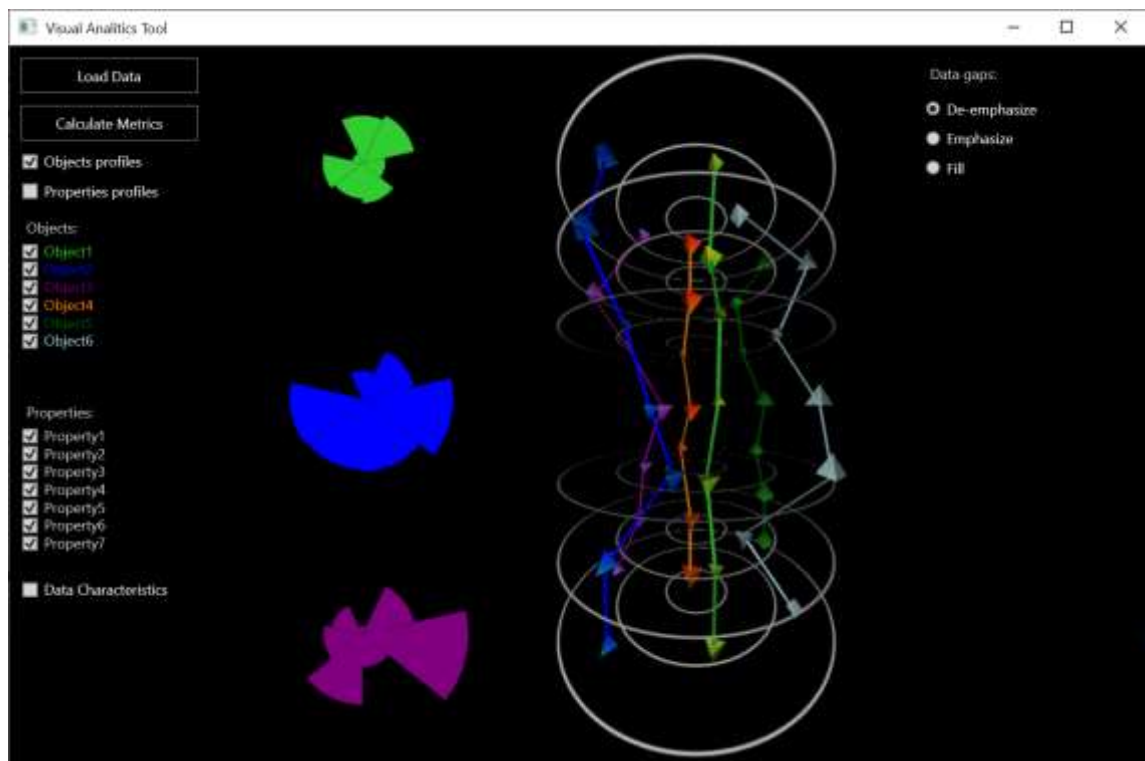


Fig. 11. The software tool's user interface

The application provides the user with the following options to manage the visual model:

- interactive filtering of data elements (objects and properties) to be visualized;
- activation, including in any combinations, of metaphors for visualization of object profiles and property profiles;
- activation of representation metaphors for visualization of various data properties (deviation of objects from the reference, deviation of object properties from average values, difference of a pair of compared objects from each other);
- selection of metaphors for visualization of gaps in data (relevant for datasets with gaps in property values).

An important feature of the software tool is the calculation of quantitative characteristics of various elements of the visual model. This provides the possibility to integrate the imple-

mented technology into the general data analysis pipeline. Examples of such characteristics are the following:

- lengths of object profiles;
- lengths (perimeters) of property profiles;
- areas of figures bounded by property profiles;
- areas of figures that visualize deviations of objects from the reference and from the average.

The software tool supports saving quantitative characteristics of the visual model for their analysis by ‘strict’ methods (e.g., using mathematical statistics methods). This can be done in order to confirm the obtained conclusions and test the hypotheses about the studied data set.

The developed application is currently undergoing state registration. In the future, it is planned to be integrated into the software platform for processing and analyzing heterogeneous data of cyber-physical systems objects functioning, created with the participation of the authors, as a subsystem of exploratory data analysis using visual models. This subsystem is expected to interact with other analytical subsystems, including for the purpose of building a unified data analysis pipeline, as well as with auxiliary subsystems responsible for the implementation of various methods of data collection, storage and preprocessing.

6. Applying the technology to the analysis of a test dataset

Let us demonstrate a number of possibilities of exploratory analysis technology. For the demonstration we will use a test synthetic dataset, the features of which will allow us to clearly show the specific capabilities of the metaphors used. The test set is represented by 9 objects, each of which is described by 7 properties. At the same time, several objects have missing values of some properties, i.e. their descriptions are incomplete.

In particular, we give examples of objectives such as:

- identification of objects with insufficient data in their description (i.e. objects for which additional collection of missing data is recommended);
- finding properties that do not carry information useful for analysis (uninformative properties);
- identification of objects that lose out to other objects in terms of the characteristics presented;
- identification of groups of similar objects;
- finding anomalous objects that are not included in the identified groups.

Visualizing the original dataset using a three-dimensional metaphor produced the visual image shown in Fig. 12 (left). At the same time, the visual image in Fig. 12 (right) allowed us to separately examine the objects with gaps in the data and evaluate these objects for the feasibility of further investigation. Thus, it can be seen that one of the objects is described by too little data, so its investigation will not lead to reliable conclusions. At the same time, the other two objects, even if there is some incomplete knowledge about them, can be considered further.

The application of the petal metaphor provides a different perspective on the data set under study (Fig. 13). The resulting visual images also reflect the presence and location of gaps in the description of objects.

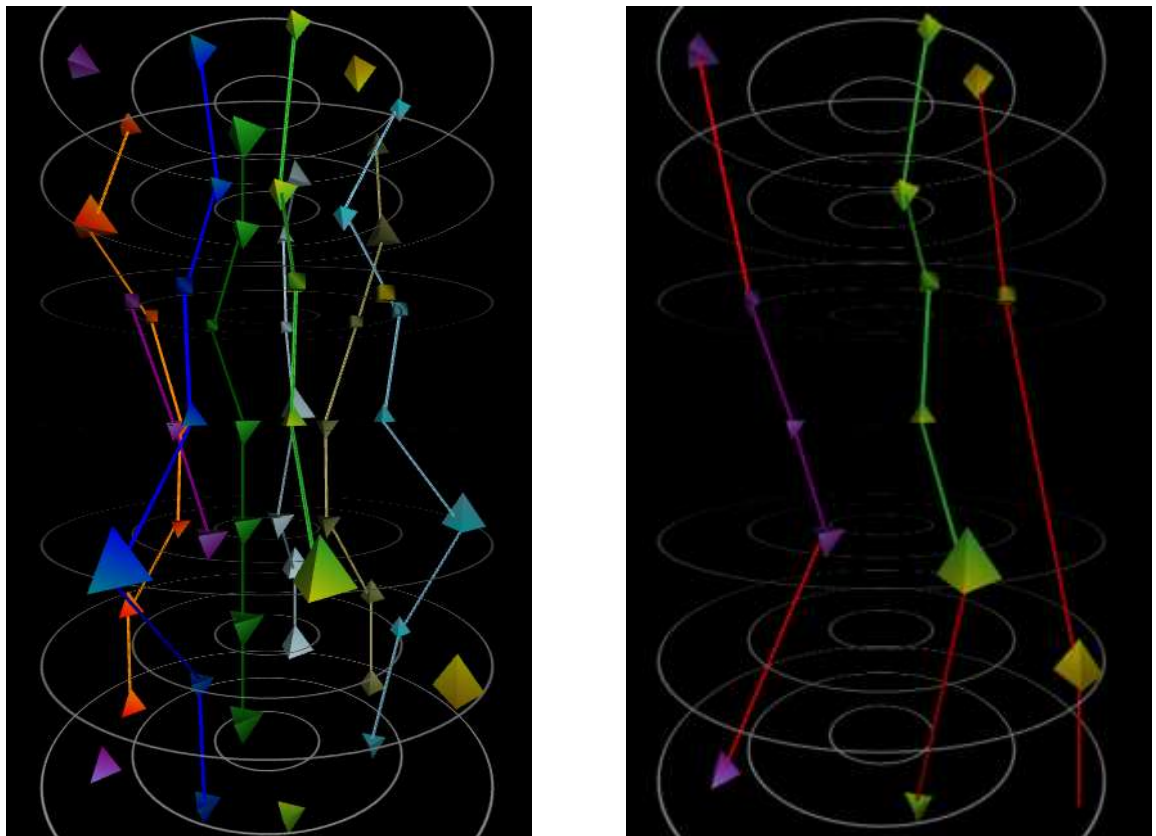


Fig. 12. Visualization of the entire dataset and visualization of objects with gaps in the data (spatial relationship metaphor)

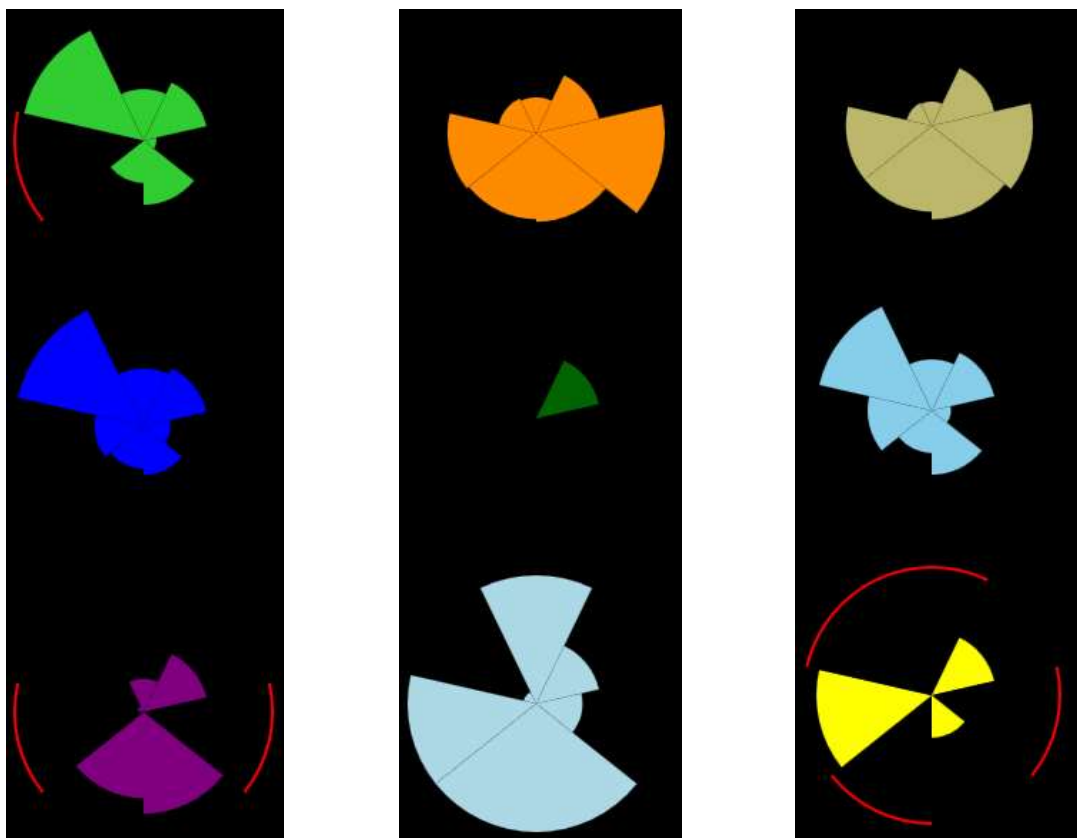


Fig. 13. Visual image of data indicating gaps in the data (petal metaphor)

The visual image in Fig. 14 was used to assess the level of informative properties of objects. A non-informative property is understood here as such a property, by which the studied ob-

jects do not differ or differ insignificantly. Thus, the profile of one of the properties has the shape of a regular polygon, which signals its uninformiveness (this is also confirmed by the type of markers on the corresponding plane).

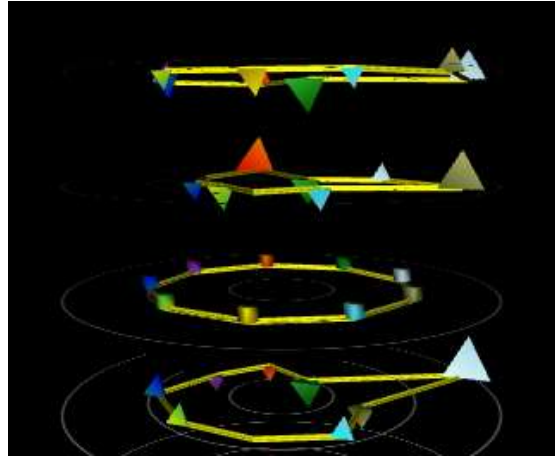


Fig. 14. Visualization of property profiles: detection of an uninformative property

After removing the uninformative property and poorly described object from the visual model, the remaining objectives were successfully achieved. The visual image in Fig. 15 (left) facilitated the detection of the object with the lowest (minimum) property values. The metaphor of deviation from the reference associated with maximum property values was utilized. Therefore, the desired object corresponds to the figure with the largest area, which can be quickly and unambiguously identified. Fig. 15 (on the right) displays an image that aids in identifying an object with the most anomalous characteristics. The figure's area represents the degree of difference between the object's property values and the average values of those properties for the entire dataset.

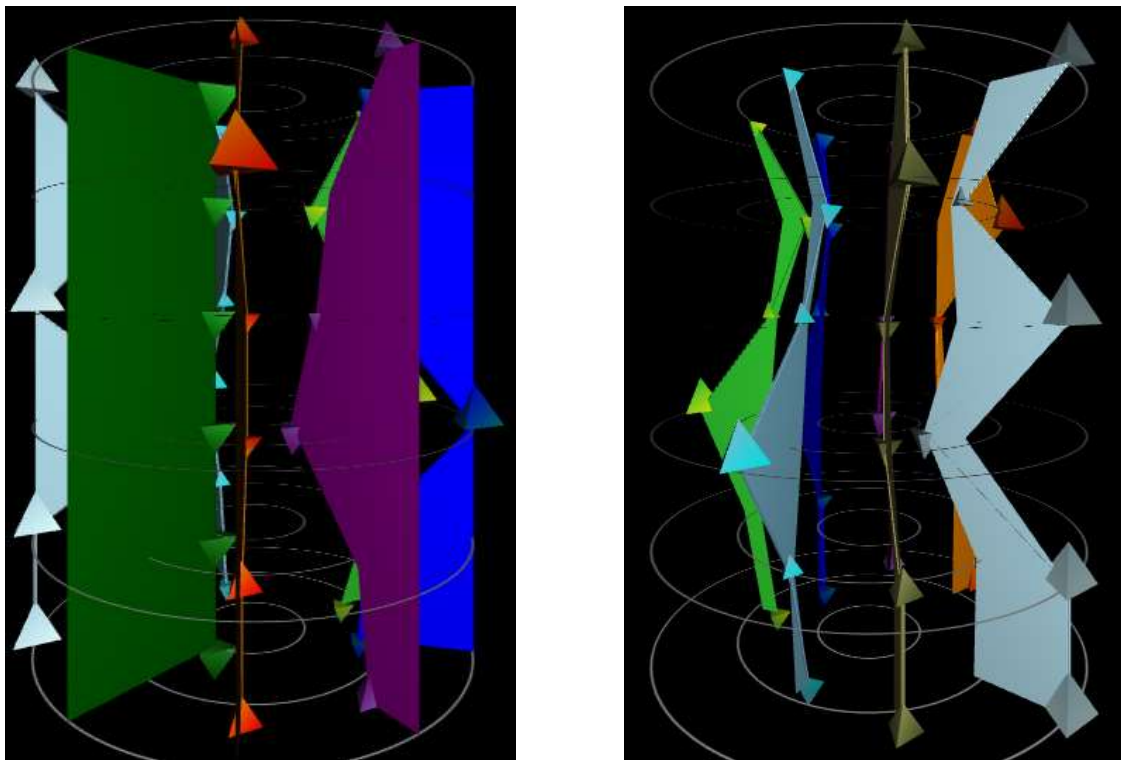


Fig. 15. Detection of objects with unsatisfactory and anomalous characteristics

Visual comparison of object profiles was used to identify groups of similar objects. Fig. 16 displays similar objects detected due to the proximity of their profiles to mirror symmetry.

The object with gaps in the description was also included in this group, as the available information justified this conclusion (Fig. 16, right). Note that if there are gaps in the description of this object that need to be filled, it is recommended to use the values of similar objects.

Another group of similar objects was found in a similar way (Fig. 17).

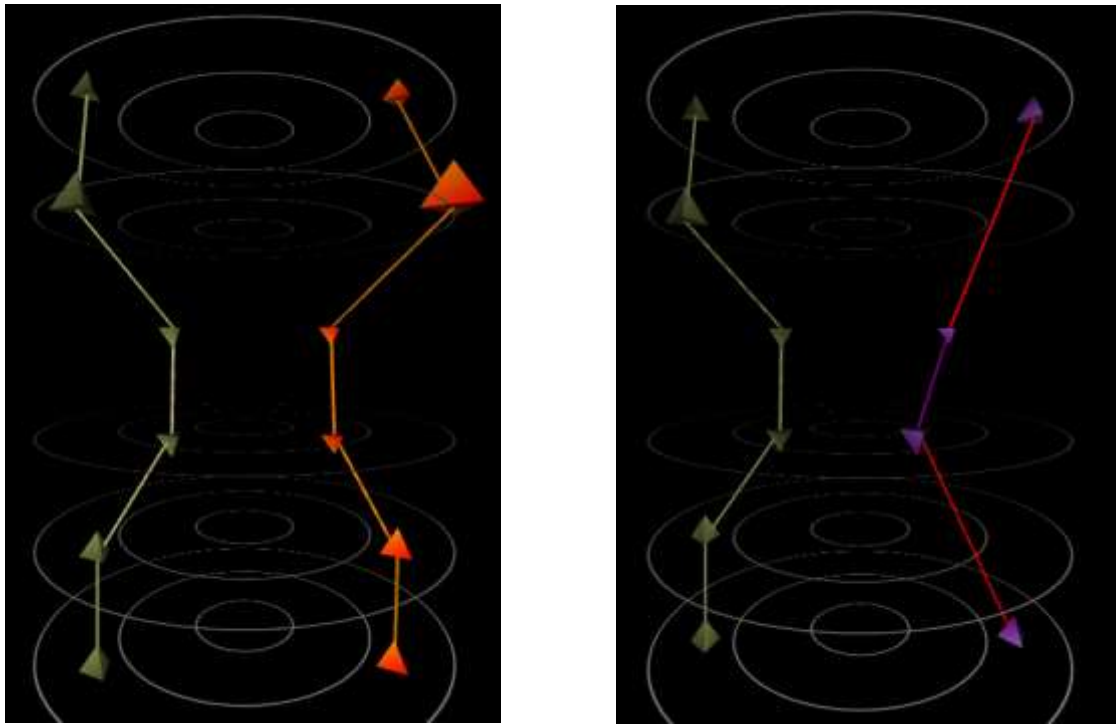


Fig. 16. Detection of similar objects, including under conditions of incomplete information

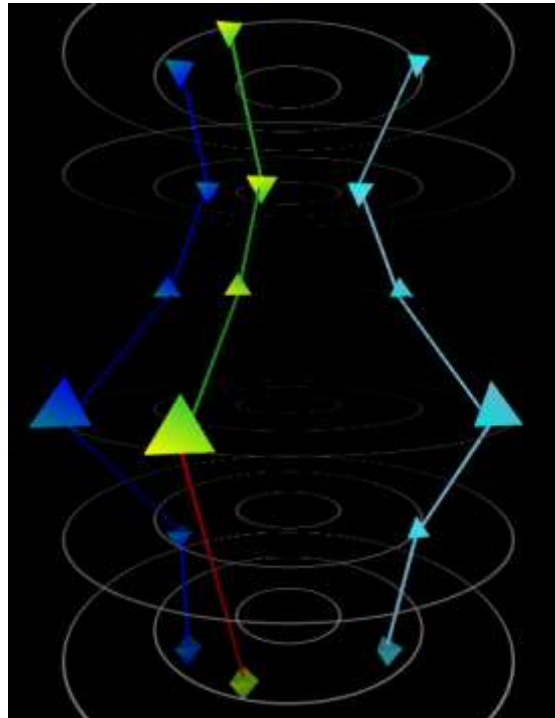


Fig. 17. Detecting another group of similar objects

Among the possibilities of interactive control, we should mention a very useful possibility of rotation of the visual model around its axis. This creates an animation that helps to quickly identify similar object profiles, select figures with the largest areas, and more accurately evaluate the shape of property profiles compared to studying a stationary model. The application

of such a capability corresponds to the inner loop (“Perception of visual image of data” – “Interactive model control”) in the technology scheme (Fig. 10).

Furthermore, the quantitative characteristics of the visual model were calculated. The values of several metrics for each element of the model are presented in Tables 1 and 2. It is noteworthy that these values corroborate the conclusions about objects and their properties that were reached during the visual analysis. In the future, these and other metrics can be subjected to analysis using the methods of mathematical statistics in order to test the hypotheses put forward about the data.

Table 1. Examples of visual model metrics (for objects)

	Object								
Metric name	1	2	3	4	5	6	7	8	9
Profile length	1.142	1.132	1.101	1.094	1.083	1.336	1.07	1.12	1.071
Area of deviation	0.507	0.517	0.583	0.425	0.917	0.34	0.492	0.523	0.5

Table 2. Examples of visual model metrics (for properties)

	Property						
Metric name	1	2	3	4	5	6	7
Property perimeter	1.34	1.441	1.384	1.355	1.026	1.546	1.236
Property area	0.418	0.514	0.393	0.29	0.435	0.293	0.469

The technology was tested not only on a synthetic data set but also on data from a real cyber-physical system. Specifically, the experiments analyzed related to unmanned aerial vehicle (UAV) flights and tested the operation of GPS during flight. Nine experiments were conducted in total, and such data as planned and actual experiment duration, UAV battery capacity, air temperature and humidity, atmospheric pressure, wind speed, etc. were collected for analysis.

The dataset was visualized using the developed software tool. The results of analyzing the visual model allowed:

- find out parameters that are not informative for analyzing flight mission models in the conditions of the experiments conducted;
- detect experiments with an abnormally high value of the actual duration, which, in turn, allowed us to detect a technical error in fixing the end time of one of the experiments;
- detect experiments described by identical sets of values, which is also due to technical errors in fixing their results.

Thus, it is confirmed that the software tool can be effectively used in exploratory analysis of cyber-physical systems functioning data.

7. Conclusion

This paper presents a technology for exploratory visual analysis of heterogeneous data, which is based on the joint application of two visualization metaphors. The paper considers the possibilities of these visualization metaphors and provides examples of visual images of data that can be obtained using the metaphors.

The general scheme of the technology is described, and it is shown that the process of visual data exploration is iterative. The text examines the semantic content of various iterative actions that an analyst performs during a study.

A software tool has been developed and tested for working with the visual model of heterogeneous data, which implements the presented technology. The tool includes an important

functionality for calculating and exporting metrics (quantitative characteristics) of the visual model. These metrics can be used for subsequent analysis by other methods to more rigorously test hypotheses made about the data.

An example of using a software tool for exploratory analysis of synthetic datasets is presented to demonstrate the technology's various aspects and visualization possibilities. The results confirm the feasibility of utilizing the technology and software tool for real-world visual data analysis tasks. The additional approbation on the experimental data with real cyber-physical system confirms the possibility of using the developed technology and software tool for exploratory analysis of cyber-physical systems functioning data.

Prospective studies can be conducted in the following related areas.

- Approbation of the presented technology on real heterogeneous data from different subject areas. This will allow to identify the limitations of the technology and metaphors for their subsequent modernization.
- Expanding the capabilities of the described metaphors by developing new representation metaphors for visualization of new indicators and data characteristics, including systemic ones. It is also intended to expand the composition of metrics (quantitative characteristics) of the visual model that can be calculated.
- Development of the proposed technology through the development of new visualization metaphors for it (both two-dimensional and three-dimensional), including for their joint use in various combinations.
- Development of ways to integrate the technology into the general data analysis pipeline, including the export of quantitative characteristics of the visual model for further analysis.

8. Acknowledgements

This study was funded by the Russian Science Foundation, project number 23-19-00342, <https://rscf.ru/en/project/23-19-00342/>.

References

1. Tukey J.W. Exploratory Data Analysis. Pearson, London, 1977.
2. Chatfield C. Exploratory data analysis. European Journal of Operational Research, 1986, Vol. 23(1), pp. 5-13. doi: 10.1016/0377-2217(86)90209-2
3. Komorowski M., Marshall D.C., Saliccioli J.D., Crutain Y. Exploratory Data Analysis. In: Secondary Analysis of Electronic Health Records. Springer, Cham, 2016. doi: 10.1007/978-3-319-43742-2_15
4. Verbeeck N., Caprioli R.M., Van de Plas R. Unsupervised machine learning for exploratory data analysis in imaging mass spectrometry. Mass Spectrometry Reviews 39(3), 245–291 (2020). doi: 10.1002/mas.21602
5. Wang G., Zhao B., Wu B., Zhang C., Liu W. Intelligent prediction of slope stability based on visual exploratory data analysis of 77 in situ cases. International Journal of Mining Science and Technology, 33(1), 47–59 (2023). doi: 10.1016/j.ijmst.2022.07.002
6. Averbukh V.L. Semiotic Approach to Forming the Theory of Computer Visualization. Scientific Visualization, 5(1), 1–25 (2013).
7. Tricoche X., Garth C. Topological Methods for Visualizing Vortical Flows. In: Möller T., Hamann B., Russell R.D. (ed.), Mathematical Foundations of Scientific Visualization, Computer Graphics, and Massive Data Exploration. Mathematics and Visualization. Springer, Berlin, Heidelberg, pp. 89-108 (2009). doi: 10.1007/b106657_5
8. Bondarev A.E., Galaktionov V.A. Generalized Computational Experiment and Visual Analysis of Multidimensional Data. Scientific Visualization, 11(4), 102–114 (2019). doi: 10.26583/sv.11.4.09

9. Galkin V.A., Dubovik A.O. Visualization of flows of a viscous conductive liquid with the presence of impurities in the flow field corresponding to exact solutions of the MHD equations. *Scientific Visualization*, 13(1), 104–123 (2021). doi: 10.26583/sv.13.1.08
10. Galkin T.P. Grigoryeva M.A., et al. An Application of Visual Analytics Methods to Cluster and Categorize Data Processing Jobs in High Energy and Nuclear Physics Experiments. *Scientific Visualization*, 10(5), 32–44 (2018). doi: 10.26583/sv.10.5.03
11. Namiot D.E., Romanov V.Yu. 3D Visualization of Architecture and Metrics of the Software. *Scientific Visualization*, 10(5), 123–139 (2018). doi: 10.26583/sv.10.5.08
12. Bondarev A.E., Bondarenko V.A., Galaktionov V.A. Visual Analysis of Text Data Volume by Frequencies of Joint Use of Nouns and Adjectives. *Scientific Visualization*, 12(4), 9–22 (2020). doi: 10.26583/sv.12.4.02
13. Zakharova A.A., Shklyar A.V. Visualization Metaphors. *Scientific Visualization*, 5(2), 16–24 (2013).
14. Podvesovskii A.G., Isaev R.A. Visualization Metaphors for Fuzzy Cognitive Maps. *Scientific Visualization*, 10(4), 13–29 (2018). doi: 10.26583/sv.10.4.02
15. Isaev R.A., Podvesovskii A.G. Cognitive Clarity of Graph Models: an Approach to Understanding the Idea and a Way to Identify Influencing Factors Based on Visual Analysis. *Scientific Visualization*, 14(4), 38–51 (2022). doi: 10.26583/sv.14.4.04
16. Zakharova A.A., Shklyar A.V. Visual Presentation of Different Types of Data by Dynamic Sign Structures. *Scientific Visualization*, 8(4), 28–37 (2016).
17. Zakharova A.A., Korostelyov D.A., Fedonin O.N. Visualization Algorithms for Multi-criteria Alternatives Filtering. *Scientific Visualization*, 11(4), 66–80 (2019). doi: 10.26583/sv.11.4.06