VISUALIZATION METHOD FOR MULTIDIMENSIONAL RANDOM PROCESSES

Oleksandr Poliarus, Dr. Sc., Prof., Andrii Lebedynskyi, PhD., Yevhenii Chepusenko, PhD Student.,
National Automobile and Highway University, Ukraine;
Nina Lyubymova, Dr. Sc., Prof.,
State Biotechnological University, Ukraine;
e-mail:yevhenii.chepusenko@gmail.com

https://doi.org/

Abstract. The article proposes a method for visualizing multidimensional random process realizations using the example of the concentrations of harmful gases emitted into the atmosphere from a thermal power plant. The method is based on the transformation of gas concentration values in one point of multidimensional space at the same time into a two-dimensional curve, which is described by the sum of products of normalized concentrations by orthogonal Legendre functions of the corresponding order. The combination of such curves on a two-dimensional plane at discrete times creates a characteristic image that can be used to visually detect features of gas concentrations over time by a human operator.

Key words: visualization, multidimensional random process, visualization method

1. Introduction

At many technical objects, a large number of parameters of heterogeneous physical quantities are simultaneously measured, which are necessary to decide the state of this object or its parts. For example, the ecological state of the environment in the area where a thermal power plant (TPP) is located depends on the concentration of harmful gases (CHG) emitted into the atmosphere. The estimation of exceeding the acceptable threshold of the CHG should be evaluated for each parameter using statistical data based on known theories [1] during the computer processing of measured data. Multidimensional processing of CHG is appropriate for creating important practice information about the state of the TPP. Currently, many methods have been developed for processing multidimensional data [2], in particular, in data mining methods [3], which have shown their effectiveness.

However, at some facilities, for example, nuclear power plants, measuring instruments are widely used for visual monitoring of equipment operation the number of which can be large. Such devices include, in particular, manometers that are highly reliable when operating under difficult conditions of high temperature, vibration, etc. The plant operator must monitor the indicators of the instruments and his fatigue can adversely affect the quality of decisions on the state of the object. Therefore, the expediency of multidimensional data visualization arises, which involves matching many random processes to one two-dimensional visual image, which will mean something to an operator of average skill. This is the task of information visualization, which can be carried out by various methods [4].

A feature of information visualization is the lack of a general theory due to the complexity of the problem, the solution of which must be carried out using both the technical characteristics of the object and measuring equipment and the biological and psychophysiological factors of the human operator. The components of this theory should be cognitive, perceptual theories, and the general theory of measurements and all these theories should use statistical methods. The interpretation of visualization results should be based on syntactic, semantic, pragmatic, and stylistic structures, as well as communication theory [5], [6]. Visualization of information is carried out during the processing of experimental data, which requires compliance with the requirements for the accuracy of measuring the corresponding physical quantities. The ultimate goal of visualization is to gain an understanding of the state of an object based on a mental model of phenomena. Since the concept of understanding is inextricably linked with a person, many mental models can exist. To avoid this, it is necessary to create a reference model of a human operator with specified physical and psychophysiological characteristics, and then the concept of "reference perception of information" is possible, similar to the concept of a standard in metrology. The transition from measuring the parameters of physical quantities to understanding is impossible without the introduction of an abstraction that allows one to conclude the characteristic features of the resulting visual image, which should be described by a pattern or patterns. Their form, as a rule, depends on the structure of the measured data and visualization tasks.

The main disadvantage of existing models is that they emphasize different parts of the communication process between the person and the obtained measurement data.

2. Goal

The purpose of the article is to develop a new visualization method for many realizations of a multidimensional random process using the example of the concentration of harmful gases emitted from a thermal power plant.

3. Information visualization methods

Consider the main methods of information visualization. In [7] an analysis of methods for creating projections of multidimensional data on two-dimensional planes was carried out, and methods for studying multivariate statistical data through animation were developed. For the perception of the complex abstract structure of the object, viewing from all sides is proposed. This leads to the creation of a sequence of two-dimensional images in the human brain. The article considers a set of realizations of many random processes that describe, in the general case, different physical quantities. For one moment, we have n values of these quantities, which n coincides with the number of random processes. A set of n values (numbers) describes some abstract object or point. The realizations of these processes for other moments are characterized by a different set of values. Let the number of time points be equal to m. Then the information about the realization of random processes can be described using n points in the m-dimensional Euclidean space. After some time, other measurements are made and a different set of process realizations is obtained, which is already described by a new point in the m-dimensional space. This space cannot be represented by the human brain, and therefore its visualization is necessary using various methods. D. Asimov offered an overview of the sequence of projections on two-dimensional planes. This creates a movie effect when any multidimensional data set (one point in m-dimensional Euclidean space) turns into a sequence of two-dimensional planes. Then the directions in these planes are selected, corresponding to the horizontal and vertical directions on the computer monitor screen. Each animation frame is an orthogonal projection of the data onto a 2D space.

In [8] attention is brought to the terminology: the term «multidimensional» refers to the dimension of independent variables, while the term «multivariate» refers to the dimension of dependent variables. In an experiment, it is not always possible to unambiguously assert the dependence of individual physical quantities. Methods for projecting multidimensional data onto a two-dimensional plane are considered in many works. An overview of such methods is given in [9] a classification of such methods is given. The main problems of big data visualization are formulated in [10]. Data visualization allows a researcher even with a low level of mathematical training to make correct decisions. In [11] an analysis of the presentation of multidimensional data in the form of a pictograph, for example, based on Chernov's faces [12]. Many illustrative examples of visualization results are provided in [13]. In [14] was introduced a new transparent approach to rendering on the web, where designers selectively bind inputs to arbitrary document elements, applying dynamic transformations to both create and modify content. In [15] the application of analytical methods and big data models for the Internet of things and visualization is considered. Visualization of multidimensional data can be based on neural networks using principal sensitivity analysis to image brain activity [16]. Authors [17] created a new t-SNE method that visualizes multidimensional data on 2D or 3D planes. Demonstration of this method on many data samples showed its advantage over other nonparametric visualization methods: Sammon mapping, Isomap, Locally Linear Embedding. To simplify the visualization, the dimension of the data can be reduced [18]. Dimension reduction methods are given in [19].

In practice, two-dimensional relationships between all pairs of variables, which can be quantitative and categorical, are often used to visually describe a multidimensional data set. A scattering matrix is a natural tool for the graphical exploration of these relationships. In [20] using a generalized matrix method that allows you to simultaneously use both quantitative and categorical variables was proposed. In [21] method was improved by introducing interactive approaches. The HyperSlice method is described in detail in [22]. Visualization of functions of two and three variables is performed by traditional methods. If there are more variables, it is necessary to fix the value of some variables so that the number of free variables is less than four, and then use the standard visualization. So, slices of multidimensional data are initially selected and visualized sequentially. The geometric coordinates denote two variables, and the gray or color value denotes the value of the function. However, one slice only shows a very limited subset of the higher-dimensional space. Therefore, the HyperSlice method was developed, in which the function is presented simply and understandably and all parameters are processed in the same way. The central concept is the representation of a multidimensional function as a matrix of two-dimensional orthogonal slices.

4. Description of the experimental research results

To explain the method of visualization of the multidimensional random process CHG, let us consider its normalized realizations obtained at the TPP. Normalization is necessary so that the concentrations are expressed in the same units and can be compared with each other. TPP gases that are emitted into the atmosphere must be controlled and not exceed the norms established by the National Standards of Ukraine. As a result of experimental measurements of CHG at the plant, an array of data was obtained on five parameters for CO, Dust, NO_x, O_2, and SO_2. A graphical representation of the results of measuring realizations of the multidimensional random process of the CHG is shown in Fig.1-5. The analysis shows that at some intervals of time, CHG exceeds the permissible values, which can be detected during visualization.

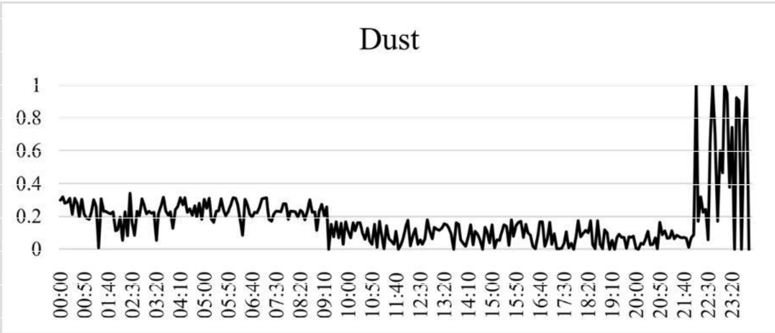


Fig. 1. Time dependence of the normalized dust concentration

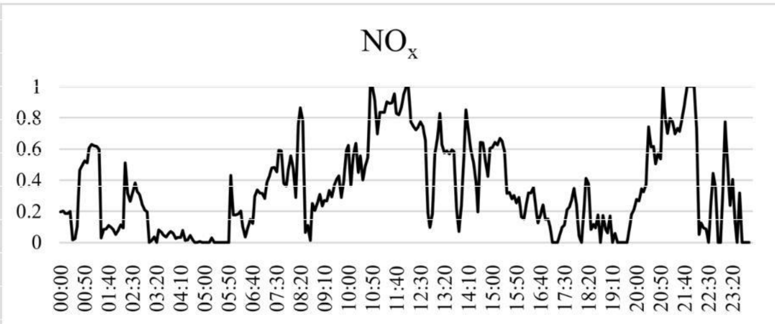


Fig. 2. Time dependence of the normalized concentration of NO_x gas

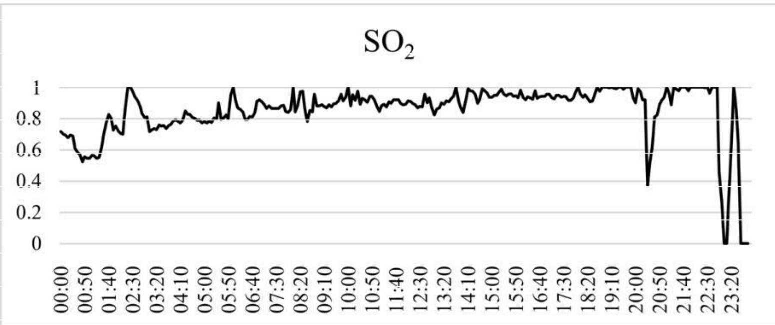


Fig. 3. Time dependence of the normalized concentration of SO_2 gas

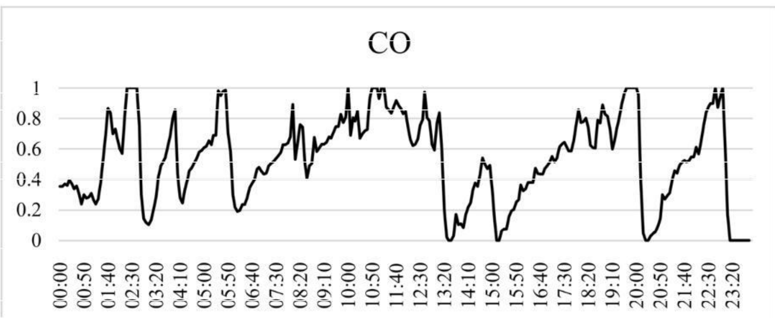


Fig. 4. Time dependence of the normalized concentration of CO gas

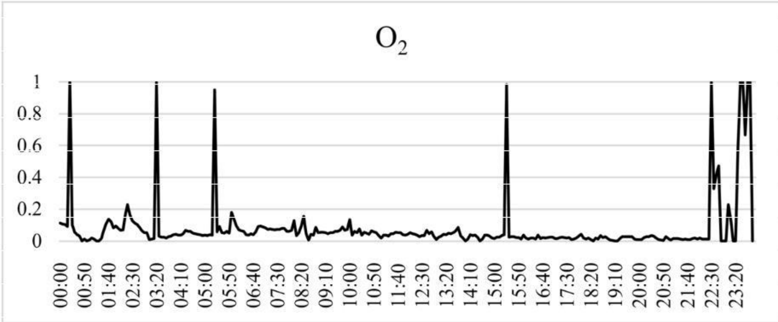


Fig. 5. Time dependence of the normalized concentration of O_2 gas

5. The essence of the proposed visualization method

Therefore, we have n realizations of normalized random processes (n = 5). Each process is discretized in time by m normalized values. The number of digital data is $n \times m$, and this number can be very large. At a discrete point in time k, we have the numerical value of the i-th measured parameter a_{ik}, i.e. in n-dimensional Euclidean space there is a A_k point that evolves in this space over time. It is necessary to follow this point visually over time to make some conclusions about the CHG emission system.

Let us denote the values of these parameters at the initial moment of time $a_1, a_2, \ldots, a_n$, where n is the number of parameters. The other indices ($i = 1, 2, \ldots, n$) refer to the parameter number. The set of these parameters can be considered as a vector with components (projections) $a_1, a_2, \ldots, a_n$. Let us consider a multidimensional vector as a two-dimensional image using a basis of orthonormal functions $\{\psi_i(\tau)\}_{i=0}^{\infty}$, which, in particular, can be orthonormal functions of Gegenbauer, Hermite, Laguerre, Legendre, Chebyshev, etc. The authors conducted a study of the visualization of random processes using different functions.

The most informative visual representations are provided by the Legendre functions $\{l_i(\tau)\}_{i=0}^{\infty}$ on an interval [0,1] where τ is a parameter that combines all the other parameters on which the state of the object depends. A similar equation for data processing in [23] was given. A point (vector) in a multidimensional space $\vec{a} = (a_1, a_2, \ldots, a_n)^T$ can be associated with the function (1).

$$\rho(\tau) = \sum_{i=0}^{n-1} a_i l_i(\tau). \qquad (1)$$

To visually analyze the behavior of a multidimensional point, you can build a three-dimensional graph $\rho(\tau, t)$ or look at the behavior of the function $\rho(\tau)$ at consecutive discrete points in time. Such a function can be a test function, that is, defined at the initial time with sufficient accuracy. Thus, each point in multidimensional space corresponds to one abstract curve of the plane. The need for abstraction is discussed in the introduction. This curve describes the set of parameters at a certain point in time. The visualization of a point in a multidimensional space cannot be represented. Moving to an abstract curve means representing a multidimensional point in two-dimensional space, which can now be easily represented.

Therefore, equation (1) includes the normalized experimental values of the harmful i-th gas concentration a_i and the corresponding orthogonal functions $l_i(\tau)$. The set of such curves for different moments creates some abstract visual image that needs to be learned to interpret. We will call it the characteristic image. The creation of such a view is the essence of the proposed visualization method. Figure 6 shows the dependencies $\rho(\tau)$ for the different order of the parameters a_i alternation in equation (1).

The order of the CHG, such as CO, dust, NO_x, O_2, SO_2 or NO_x, O_2, dust, SO_2, CO, affects the view of the image. So that the conclusion regarding the state of CHG does not depend on the order of recording the concentrations of these gases, one must choose any order and then always use it.

Now let's consider the influence of the jumps of the CHG on the image view. To do this, we artificially exclude jumps in dust and O_2 concentrations from the results of experimental studies. Figure (7, a) shows an abstract image of CHG with jumps of the indicated concentrations, and Figure (7, b) shows the same picture, but with the extracted jumps.

From Fig.7 follows that the characteristic image changes after the jumps are extracted: the lower curves disappear (Fig. 7, a), which describes the jumps of the CHG in the abstract space. Next, the statistics of such images with different jumps exceeding the DSTU thresholds and without jumps are collected. This may be the subject of a separate study that will help create a characteristic image pattern. Exceeding its limits will indicate the presence of anomalous emission of the concentration of one of the harmful gases, the name of which can be determined by a specially developed algorithm similar to that developed in [24]. One pattern, for example, a rectangle, an ellipse, etc., is superimposed on the obtained characteristic images (Fig. 6, 7). A change in the size of the pattern means a different probability of CHG jumps detection for a given value of the probability of a false alarm.

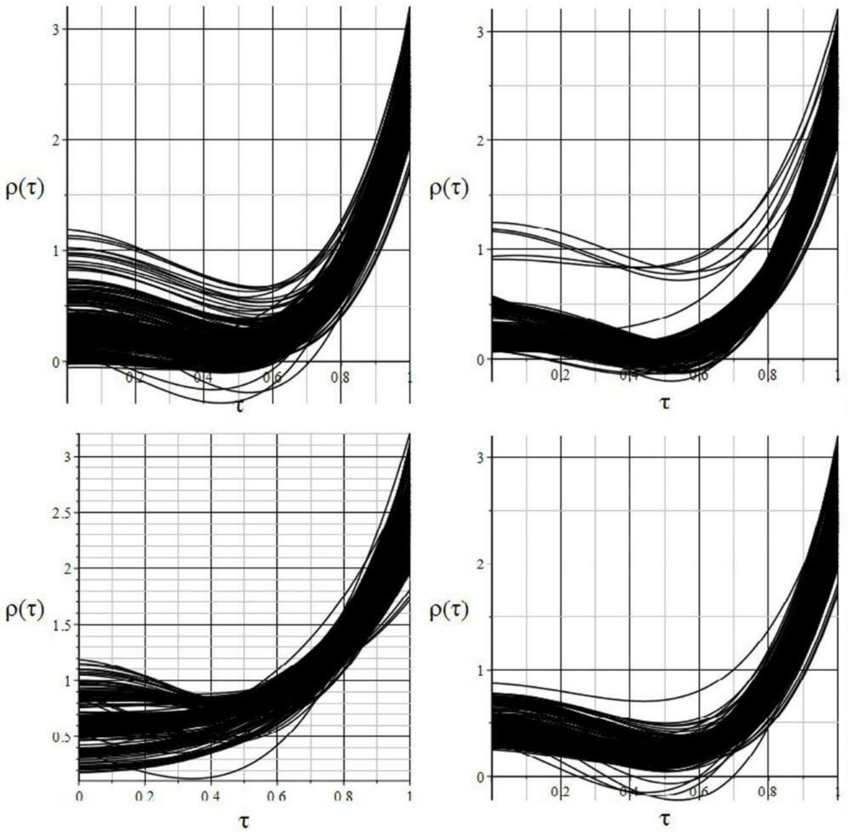
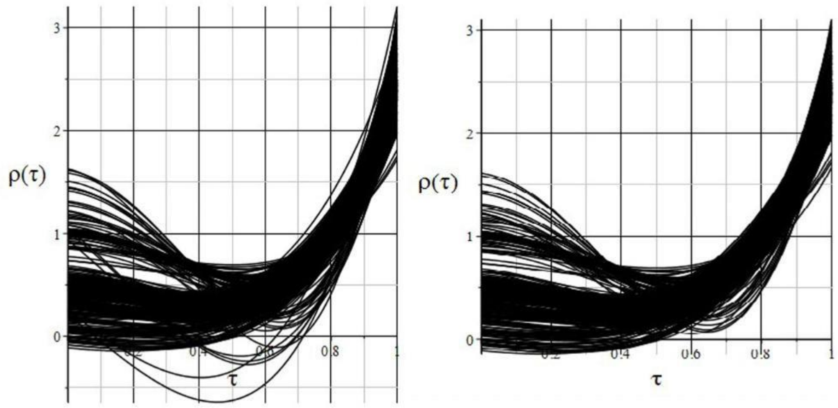


Fig. 6. Visual image of the CHG using the Legendre functions.



a *b*

Fig. 7. The results of the jumps impact on the visualization of the CHG: a – with jumps: b – without jumps.

6. Conclusions

Most of the visualization methods are aimed at describing the local properties of the data. For the estimation of general properties, the article proposes a method for visualizing multidimensional random processes using the example of the concentration of harmful gases, the number of which can be large. To obtain a characteristic image, it is enough to have a set of data that characterizes the results of measuring individual parameters of random processes, and orthogonal functions. So, the article solves the problem of visualizing a multidimensional process that describes the behavior of various harmful gas concentrations.

7. Gratitude

The authors thank the teachers of the department of metrology and life safety of the Kharkiv National Automobile and Highway University for discussing the results of the study and for their support.

8. Mutual claims of authors

The authors have no claims against each other.

References

[1] G. Phillips-Wren. Intelligent Systems to Support Human Decision Making. In book: Artificial Intelligence, 2017, pp. 3023–3036. DOI:10.4018/978-1-5225-1759-7.ch125

[2] S. Mansmann, T. Neumuth, M. H. Scholl, Multidimensional Data Modeling for Business Process Analysis, 26th Int. Conf. on Conceptual Modeling, Nov. 5-9, 2007, Auckland, New Zealand. DOI:10.1007/978-3-540-75563-0_4

[3] J. Starck, F. Murtagh, *Handbook of Astronomical Data Analysis.* Elsevier, 2002. [Online] Available: https://www. academia.edu/2608657/Handbook_of_Astronomical_Data_Analys is

[4] Pak Chung Wong and R. Daniel Bergeron. 30 Years of Multidimensional Multivariate Visualization. In *Sc. Visualization, Overviews, Methodologies and Techniques. IEEE Computer Society Press*, pp 3–33, 1994. Available: https:// www. cs.unc.edu/xcms/courses/comp715-s10/papers/Wong97_30_ years_of_multidimensional_multivariate_visualization.pdf

[5] H. C. Purchase, N. Andrienko, T. J. Jankun-Kelly, M. Ward. Theoretical Foundations of Information Visualization. In: Inf. Visualization: Human-Centered Issues and Perspectives, 1970, pp.46-64. DOI:10.1007/978-3-540-70956-5_3.

[6] W. Weaver, C. Shannon. The mathematical theory of communication. Physics, 2009. DOI:10.1098/rspa.2009.0063

[7] D. Asimov, “The grand tour: A tool for viewing multidimensional data”, SIAM Journ. on Sc. & Stat. Comp., pp. 128-143, 1985. DOI: 10.1137/0906011

[8] R. Bergeron, W. Cody, W. Hibbard, D. Kao, K. Miceli, L. Treinish, S. Walther. Database Issues for Data Visualization: Data Model Development. In IEEE Visualization '93 Workshop, San Jose, California, USA, October 26, 1993, pp. 3-15. *In Proc. Lecture Notes in Comp. Sc.ce 871*, Springer 1993. [Online] Available: https://link.springer. com/book/10.1007/BFb0021138

[9] I. Romanova, "Modern Methods of Multidimensional Data Visualization: Analysis, Classification, Implementation and Applications in Technical Systems, Science and Education of the Bauman MSTU, Vol. 3, 2016, pp. 133–167. DOI: 10.7463/0316.0834876

[10] Zongben Xu, Yong Shi. Exploring Big Data Analysis: Fundamental Scientific Problems, Ann. data sci., 2 (4), 2015, pp. 363-372. DOI:10.1007/s 40745-015-00637

[11] Yau Nathan. Visualize This: The Flowing Data Guide to Design, Visualization, and Statistics s. Indianapolis, In: Wiley Publishing, 2011. [E-book] Available: https://www.perlego.com/book/1011299/visualize-this-the-flowingdata-guide-to-design-visualization-and-statistics-pdf

[12] H. Chernoff. The Use of Faces to Represent Points in K-Dimensional Space Graphically. Journ. Am. Stat. Ass., Vol. 68, No. 342., pp. 361-368, 1973. DOI:10.2307/2284077

[13] J. Heer, M. Bostock, V. Ogievetsky, A Tour through the Visualization Zoo. A survey of powerful visualization techniques, from the obvious to the obscure. Communications of the ACM. Stanford University. 2010. Vol. 53, Iss.6, pp.59-67. DOI:10.1145/1743546.1743567

[14] V. Ogievetsky, J. Heer. D3: Data Driven Documents, IEEE Trans. Visualization & Comp. Graphics, 2011. [Online], Available: http://vis.stanford.edu/files/2011-D3-InfoVis.pdf

[15] M. Marjani, F. Nasaruddin, A. Gani, A. Karim, I. Abaker, T Hashen, A Siddiqa, I. Yaqoob. Big Data Analytics: Architecture, Opportunities, and Open Res. Challenges. IEEE Access, vol. 5, 2017, pp. 5247–5261. DOI:10.1109/ ACCESS.2017.2689040

[16] S. Koyamada, Y. Shikauchi, K. Nakae, M. Koyama, S. Ishii. Deep Learning of FMRI big data: a novel approach to subject-transfer decoding. – arXiv: 1502.00093v1 [stat ML] 31 January 2015. [Online]. Available: https://arxiv.org/ pdf/1502.00093.pdf

[17] L. van der Maaten, G. Hinton. Visualizing Data using t-SNE. Journ. Mach. Learn. Res., 2008, vol. 9, pp.2579–2605. [Online]. Available: https://www.jmlr.org/papers/ volume9/vandermaaten08a/vandermaaten08a.pdf

[18] A. Genender-Feltheimer. Visualizing High Dimensional and Big Data. Complex Adaptive Systems Conference with Theme: Cyber Physical Systems and Deep Learning, CAS, 2018, 5-7 Nov. 2018, Chicago, USA, pp.112–121. DOI: 10.1016/j.procs.2018.10.308

[19] K. Börner, C. Chen, K. Boyack. Visualizing knowledge domains. An. Rev. of Inf. Sc. & Techn., vol. 37, 2003, Medford, NJ: Information Today, Inc./Amer. Soc. for Inf. Sc. and Techn., Ch.5, pp.179–255. DOI:10.1002/ aris.1440370106

[20] J. Emerson, W. Green, B. Schloerke, J. Crowley, D. Cook, H. Hofmann, H. Wickham. The Generalized Pairs Plot. Journ Comp. and Graph. Statistics, vol. 22(1), 2013, pp. 79-91. DOI: 10.1080/10618600.2012.694762

[21] J. Im, M. McGuffin, R. Leung. GPLOM: Generalized Plot Matrix for Visualizing Multidimensional Multivariate Data, IEEE Trans. on Visualization and Comp. Graphics, 19 (12), 2013, pp. 2606-2614. DOI: 10.1109/TVCG.2013.160

[22] J. van Wijk, R. van Liere. HyperSlice: Visualization of Scalar Functions of Many Variables, 1998. [Online]. Available: www.researchgate.net/publication/2660434_ HyperSlice

[23] D. Andrews. Plots of high-dimensional data, Biometrics, Vol. 28, no.1, 1972, pp. 69-97. DOI:10.2307/2528964

[24] O. Poliarus, Y. Poliakov, A. Lebedynskyi. Detection of landmarks by autonomous mobile robots using camera-based sensors in outdoor environments. *IEEE Sensors Journal*, vol. 21, iss.10, 2021, pp. 11443-11450, DOI:10.1109/JSEN.2020.3010883