

# New Method of Data Mining in Practical Cosmology

© Vladimir Vitkovskiy<sup>1</sup> and Vladimir Gorohov<sup>2</sup>

<sup>1</sup> Special Astrophysical Observatory of RAS, Nizhnij Arkhyz, Russia

<sup>2</sup> ENGECON State University, Sankt-Peterburg, Russia

**Abstract.** There are two kinds of computer graphics: the illustrative one and the cognitive one. Appropriate the cognitive pictures not only make evident and clear the sense of complex and difficult scientific concepts, but promote, - and not so very rarely, - a birth of a new knowledge. On the basis of the cognitive graphics concept, we worked out the Space Walker (SW)-system for visualization and analysis. It allows to train and to aggravate intuition of researcher, to raise his interest and motivation to the creative, scientific cognition, and to discover new relations in multidimensional spaces of observable parameters.

The modern observational cosmology brings more and more volumes of experimental data. Only the astrophysical observation archives have up to petabytes these now. The enormous data analysis is coming to a huge problem for investigators. We offer a rather natural way for resolution of this complex of difficult scientific and problems. These problems can be resolved with the help of modern information technologies and, first of all, with the help of the cognitive computer graphics. The SW project aim is the creation of concept and methodological means of the pictorial, descriptive visualization of the data containing in the multidimensional catalogs and the data bases. In the project resolving the fundamental problems of the use of the methodology of experimental sciences developed in phenomenology is assumed in connection with the data reduction processes. New methodology must ensure the successful application of software of the visualization of multidimensional data and systems of the visual programming. New procedures and means of work with the cognitive graphic figures give impetus to the development of fundamentally new algorithmic, software of the visualization of experimental data.

The scientific value of work consists in the determination of laws governing the formation of cognitive means in the consciousness of investigator for developing the procedures of manipulation with the cognitive means in the contemporary systems of machine drawing.

## **The project objectives:**

1. Development of procedures and methodology of work in the contemporary systems of the cognitive machine drawing, which ensure the application of these systems with the analysis of the large volume of the experimental data of the obtained in the natural sciences, especially in astronomy and astrophysics.
2. Development of the philosophical aspects of the cognitive processes of cognitive machine drawing.
3. Formation of ergonomic recommendations and requirements for software of the making decisions and visual programming support, those use the developed procedures and methodologies.
4. Resolving the fundamental problems of the use of the methodology of experimental sciences developed in phenomenology is assumed in connection with the data reduction processes. The approach to the use of methods of phenomenology proposed in the project is fundamentally new.
5. Creation of algorithms and software of the cognitive visualization of the multidimensional massifs of data.

The creation of software for the solution of the problems enumerated above in turn requires the practical mastery of the entire complex of achievements in the field of mathematical statistics, theory of illegible sets, cognitive machine drawing, cognitive psychology and, first of all, the theory of knowledge (gnoseology). Project base on the ground of possibility use already acting software intellectual support adopted solution in task control complex system with deep a priori uncertainty. This, first of all the systems, which use a cognitive machine drawing (systems SWOT, PEST- analysis, the systems of the dynamic visualization of multidimensional data of X- Gobi, SW, SCV and so forth). For overcoming the a priori uncertainty in the applied information theory the entire spectrum of technical, algorithmic, program and systematic equipment is developed. This is the first of all robust and nonparametric methods of statistical analysis, control and checking of representativeness and

uniformity of data.

The expansion of these methods in the multidimensional case presents definite difficulties. Therefore for overcoming these a priori difficulties expediently to additionally turn to the possibilities of the human intuition (is, in view, first of all professional experience, empirical knowledge on which is based a professional intuition), which makes it possible to sometimes check the quality of the work of the decisive procedures and to establish the authenticity of utilized data, which makes it possible.. It is known that the human intuition is activated in the case of the visualization of data. For the first time this was clearly demonstrated in the works Zenkin (1991), in which, in particular, it is shown as the observation of the cognitive means, connected with the properties of the natural numbers series, it made it possible to formulate and to prove a number of new theorems on the Warring problem.. At present questions of "obtaining", fixations, "mastering" of empirical experience are considered in such promising trend of information theory as engineering of knowledge. Specifically, in this direction actively are used the achievements of cognitive psychology, epistemology and mathematical means of relational algebra. Engineering of knowledge successfully demonstrated that for attraction and stimulation of intuitive knowledge undoubtedly is required new technical equipment. And such technical equipments are already created and are created. This is, first of all the system of cognitive machine drawing and system of virtual reality.

Of course experimental psychology has already sufficiently long ago developed and are investigated the effects of suggestion, and so the graphic, acoustic and tactile means of the stimulation of the consciousness of the human operator. In this sense the systems of cognitive machine drawing are only one additional subset of the interesting technical equipment, which generate new psychological phenomena. The phenomenon of cognitive machine drawing, for the first time scientifically noted by Zenkin, consists of the generation on the screen of display of the special graphic representations, which create in the brain of human operator entertainment means. These means seem man by aesthetically attractive and, thus, they stimulate its descriptive imagination, closely related to the intuitive mechanisms of thinking.

In turn intuitive descriptive thinking, being activated it generates the entire cascades of the creative scientific ideas of those facilitating both to the construction of the new scientific inductive- deductive diagrams (scientific hypotheses) and the nontrivial solutions by the organization of scientific experiments. Such algorithms can come out as basis for the new class of the scientific systems (technognostic systems), which ensure the stimulation of the scientific intuition of necessary for overcoming the problems of experimental natural science enumerated above.

We have developed concrete algorithm (Gorohov, Vitkovskiy, Ivanov 1992,1993) that capable to create on the computer screen the images, which generate in the brain of man (in his consciousness) the entertainment means, which cause strong aesthetical experiences activating, in turn, scientific intuition. This is one of the simplest algorithms of dynamic projection from the multidimensional space of data to the two-dimensional plane of visualization. The possibility of the generation of the visual means of the multidimensional cloud of data is very important, since man cannot, examining the multidimensional tables of data see these multidimensional means.

The algorithms and the programs, which achieve a projection of this cloud to the two-dimensional hyperplane  $Q$ , which coincides with the plane of the display screen, are proposed for this. The organization of projection to plane  $Q$  lies in the fact that is constructed the coordinate system, attached to this plane. A change of orienting the plane  $Q$  is expressed as the values of the guides of cosines. Thus, it is possible to examine on the screen the projection of multidimensional cloud from any possible direction in the multidimensional space it is indicative. Moreover, it is possible to set it into the cyclic rotary motion, whose direction is assigned by any vector, which lies at plane  $Q$ .

The essence of cognitive effect lies in the fact that man receives the moving projection as pseudo-three-dimensional object characterizing multidimensional means in the multidimensional space. Taking into account the fact that the user can easily change the direction of projection, its emotional possibilities can be easily begun to operate for the formation of cognitive multidimensional means. Preliminary scaling is indicative along the axes with the aid of the ordinal statistician it ensures authenticity and robustness of this means (Gorohov, Vitkovskiy 1994).. This projection makes it possible to observe the external geometric (and, therefore, statistical) properties of multidimensional means but, however, user needs the study of the internal structure of multidimensional means. Here there is the possibility to assign the hyperplanes moving in the multidimensional space, which intercept the necessary section of multidimensional space. It is possible to speak about the methods

of the straight tomography of multidimensional means. After the thorough qualitative study of the visual aspects of multidimensional means with the aid of the enumerated algorithms appears the possibility, using algorithms of standard machine drawing to paint the interesting user separate objects or the groups of objects. Then it is possible to again return to the dynamic behavior of the rotation of means for the purpose of checking the intuitive ideas of user about the clusters and the connections in multidimensional data.

Is possible the development of the methods of cognitive machine drawing in combination with other information technologies, first of all with the packets of digital processing of images and multidimensional statistical analysis. Algorithms of such type are capable of objectively reflecting those analytical properties of multidimensional data, which then are revealed by the traditional methods of multivariate analysis. In this case it is possible to grasp the new unexpected statistical connections and clusters in multidimensional data precisely for researcher, armed by the data by algorithm.. Since, here is succeeded in connecting, that resource of intuition and nonformal experience, which is characteristic precisely of man to researcher. Especially this is important with the analysis of statistical connections, the factor analysis and with pattern recognition.

The fact that the procedure of dynamic visualization does not rest on the a priori information about nature of objects, but means and it does not introduce in the projection of the distorting influences of one or other model or another, gives the possibility to use the visualized means under the conditions of deep a priori uncertainty.

The development of similar algorithms of those generating cognitive means is not at present only. The algorithms, which possess cognitive properties, are proposed by the groups of Norwegian, Finnish and American researchers. A priori uncertainty with respect to the form of the distribution of data was serious difficulty by this method. In this work it is removed due to the use of rank statistic for the scaling of data along the axes. It is known that the passage from the true measurements to order statistics ensures the independence of statistical properties from the initial distributions. Furthermore, the use of ordinal statistician ensures use as the given random- censored values also of the upper limits.

The unexpected cognitive properties of the algorithm of dynamic projection were the important aspect of this development. The researcher easily perceives this means; he easily with it begins to manipulate and freely to interpret in the terms of its specialty. Thus is destroyed the wall of alienation between the multidimensional properties of data of monitoring and the man by researcher. The cognitive shapes begin to be considered by researcher as accessible to human perception object from his subject area. It is difficult to overestimate the prospects for this phenomenon. The possibility to create the visual means of the structures of data and concepts, by which the concrete scientific experimenter or theorist manipulate appears.

A strict logical structure of these mechanisms of visualization provides the possibility of mathematical agreement and checking different theoretical diagrams. The finally fundamental logical structure of construction both the theoretical concepts and bases of experimental data will make it possible to improve the mechanisms of logical inductive conclusion relying on the nonclassical many-valued of logic. The formation of such logical structures will rest from one side to strict logical axiomatic, and from other side to experience and practice engineering and technology of scientific experiment. The joining of the intuitive ideas of theorists, experimenters and practitioners of technologists is here possible. Joining can be achieved at the level of the cognitive graphic means of these concepts and semantics corresponding to them. This, apparently, it is connected with the fact that here are used the concepts of the nonparametric statistics, which is in turn based on the theory of invariants and the probability measures. Probability measures, possibly, to treat as the formal mathematical formulation of probabilistic logician.

The development of cognitive machine drawing for the generation of the visual means of the content of the contemporary bases of given, system archives and the data banks is possible. One additional possibility of cognitive control and protection from the unsanctioned access to the content of these data bases here appears. As studies showed, even small changes in the content of the data bases sharply change their cognitive means. That also makes possible rapidly for man-operator to instantly note a change in the means of the data base, and consequently the fact of the unsanctioned access. Let us emphasize that in the form of cognitive means the content of terabyte multidimensional massifs can be represented. Furthermore, the use of the enumerated applications of a cognitive drawing can be even more effective with its introduction into the net technologies.

## References

Zenkin A. 1991, Cognitive computer graphics (Moscow: Nauka)

- Gorohov, V., Vitkovskiy, V., Ivanov, L. 1992, Robust-cognitive technology of dynamic visualisation multidimensional observational data (Nizhnij Arkhyz: SAO RAS)
- Gorohov, V., Vitkovskiy, V., Ivanov, L. 1993, Cognitive informational technology of planning and control of ecological monitoring with further robust analysis of extreme manifestations (Proceedings First Eurasian symposium on space science and technologies. Gebze, Turkey)
- Gorohov, V., Vitkovskiy, V., 1994, Algorithmic methods and software for non-parametric estimation of distribution of non-uniform observational data (Turkish Journal of Physics vol. 18, numb. 9, Published by the Scientific and Technical Research Council Turkey)