

## **Визуальная аналитика и её использование в деятельности лаборатории «Научная визуализация» НИЯУ МИФИ**

В.В. Пилюгин<sup>1,А</sup>, И. Мильман<sup>2,В</sup>

<sup>А</sup> Национальный Исследовательский Ядерный Университет «МИФИ»,  
Москва, Россия

<sup>В</sup> Самостоятельный исследователь, США

<sup>1</sup> ORCID: 0000-0001-8648-1690, [VVPilyugin@mephi.ru](mailto:VVPilyugin@mephi.ru)

<sup>2</sup> ORCID: 0000-0001-9705-9401, [Igal.Milman@gmail.com](mailto:Igal.Milman@gmail.com)

### **Аннотация**

В статье рассматривается визуальная аналитика, под которой согласно Дж. Томасу, понимается решение задач анализа данных с использованием способствующего интерактивного визуального интерфейса. На сегодняшний день визуальная аналитика широко используется в различных областях — в научных исследованиях, проектировании, управлении и других, что обусловлено известной предрасположенностью людей к пространственно-образному мышлению. Несмотря на это, как считают авторы, теоретические аспекты визуальной аналитики в настоящее время проработаны недостаточно, что безусловно является сдерживающим фактором при разработке инструментальных средств визуальной аналитики и ее эффективного использования на практике. В статье утверждается, что одной из наиболее распространенных форм визуальной аналитики является решение задач анализа данных методом визуализации. Подробно описан метод визуализации, а также приводятся примеры использования метода визуализации в ряде работ, выполненных лабораторией «Научная визуализация» НИЯУ МИФИ в этой области. Работы выполнялись в разных предметных областях от физики до финансов и мониторинга.

**Ключевые слова:** Визуальная аналитика, анализ данных, метод визуализации, моделирование.

### **1. Визуальная аналитика**

Термин «Визуальная аналитика» был введен Джимом Томасом [1] [2]. Он определил визуальную аналитику, как решение задач анализа данных с использованием способствующего интерактивного визуального интерфейса. Этот же термин используется и для обозначения научной дисциплины об этой деятельности.

В настоящее время визуальная аналитика широко используется в различных областях человеческой деятельности - в научных исследованиях, проектировании, управлении, финансовом мониторинге [3], информационной безопасности [4] и других областях. Ключ к широкому и успешному использованию визуальной аналитики в различных областях и обстоятельствах - это известная предрасположенность людей к пространственному мышлению и распознаванию образов. [5]

Перед рассмотрением Визуальной аналитики, рассмотрим проблему анализа данных в целом. Определим проблему анализа данных следующим образом:

- Возьмем объект рассмотрения  $Ob$ .
- Аналитик знает некоторые начальные утверждения об объекте  $Ob$ . Эти утверждения называются «исходными данными»  $G = \{G_i\}$ , и они могут быть формализованы в том или ином виде.
- Необходимо получить новые утверждения, представляющие интерес для аналитика или проверить данные утверждения на истинность.

В приведенном выше определении объект рассмотрения  $Ob$  может быть представлен в виде:

- Один или несколько реальных материальных объектов  $Ob = \{r_i\}$ .
- Один или несколько воображаемых материальных объектов  $Ob = \{m_i\}$ .
- Один или несколько абстрактных объектов  $Ob = \{s_i\}$ .

Как следует из приведенного выше определения, визуальная аналитика является очень широкой концепцией, и сегодня мы можем говорить о различных формах визуальной аналитики. С нашей точки зрения, одной из наиболее распространенных форм визуальной аналитики является решение задач анализа различных данных с использованием метода визуализации.

## 2. Метод визуализации

Решение задачи анализа исходных данных состоит из последовательного решения следующих двух задач (рис. 1): [6]



Рис. 1. Анализ данных методом визуализации

Первой проблемой является задача получения представления анализируемых данных в виде графического изображения (задача визуализации исходных данных). Задача решается с использованием компьютера. Вторая задача (задача получения интересующей аналитика информации), которая является не менее важной, — это визуальный анализ полученного изображения. Аналитик должен работать над этой задачей напрямую, она не может быть автоматизирована.

### 2.1. Визуализация исходных данных

Решение этой задачи включает в себя хорошо известную последовательность шагов. Алгоритм преобразования исходных данных в графическое изображение называется конвейером визуализации [7]. Шагами алгоритма являются:

- **Sourcing** — это этап получения необработанных данных  $G$ . Этот процесс может включать в себя объединение данных из разных источников или просто формализацию данных в компьютерное представление.

- **Фильтрация** — это этап предварительной обработки данных, их очистки и выполнения различных вычислительных задач для получения уточненных данных  $G^f$ . Этот шаг не является обязательным и в некоторых случаях может быть опущен.
- **Mapping** — это шаг получения математической модели пространственной сцены  $S$  из отфильтрованных данных  $G^f$ . Пространственная сцена — это набор пространственных (обычно двух или трехмерных) геометрических объектов  $G^S$  с соответствующими графическими атрибутами  $T$ . Этот шаг часто уникален для каждой задачи анализа данных. Мэппинг определяет эффективность метода визуализации.
- **Rendering** — на этом этапе на основе математической модели пространственной сцены строится ее проекционное графическое изображение  $G^r$ . Эффективность рендеринга обычно определяется качеством предыдущего шага.

Простой конвейер показан на рис. 2. Данные из файла табличного вида считываются в память компьютера в виде таблицы, фильтруются, а затем преобразуются в уравнение с помощью линейной интерполяции. Уравнение визуализируется как обычный график  $Y(X)$ .

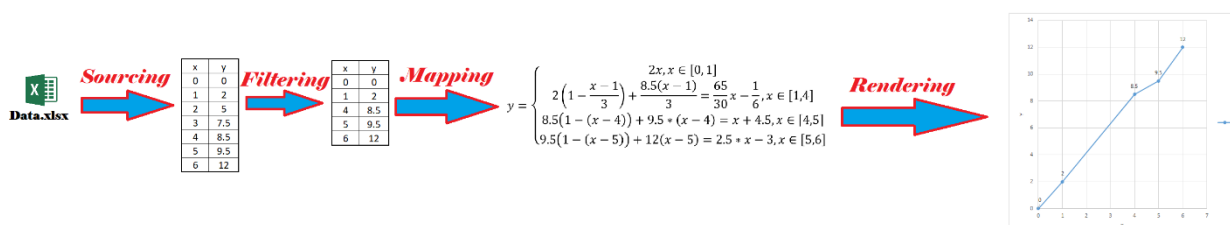


Рис. 2. Простой пример конвейера визуализации.

После завершения первой задачи, ход переходит к аналитику, которому необходимо просмотреть графические изображения и сделать интересные его утверждения.

## 2.2. Визуальный анализ графического изображения

Как указано выше, эта задача решается непосредственно аналитиком. Результаты могут и будут варьироваться в зависимости от аналитика, его образования, опыта и целей [8]. Решение этой задачи состоит из двух общих этапов:

- Визуальный анализ графического представления. Принципиально важно понимать, что полученные графические изображения являются лишь естественным и удобным средством представления пространственной интерпретации исходных данных для аналитика. Пространственная сцена должна быть визуально проанализирована. Это позволяет аналитику использовать огромные потенциальные возможности пространственного мышления в процессе анализа. На этом этапе аналитик смотрит на представление и замечает некоторые закономерности или отклонения. Этот шаг не может быть строго формализован, но можно определить основные направления анализа. Аналитик может делать следующие виды утверждений:
  - О формах объектов на пространственной сцене. Примером такого рода утверждений является «Объект № 1 — это куб, а объект № 2 — это сфера».
  - Об относительном расположении объектов. Существуют два основных способа делать подобные утверждения: 1) «объекты с первого по шестой находятся близко друг к другу» или «объект под номером семь находится далеко от всех остальных объектов». 2) «объекты с первого по восьмой расположены на гранях куба».

- Об оптических параметрах (например, цвет, прозрачность и т. д.). Примером является «Объект под номером один красный» или «Второй объект непрозрачный».

Результатом данного шага являются утверждения о визуальном представлении и объектах на пространственной сцене. Эти утверждения не имеют ничего общего с исходными данными. Важно подчеркнуть, что результаты этого шага будут отличаться в зависимости от аналитика и его способности к пространственному мышлению, обращению внимания на детали, фокус и т.д.

- Интерпретация результатов по отношению к исходным данным. На этом этапе аналитик преобразует утверждения о пространственной сцене в утверждения об объекте рассмотрения *Ob*. Этот шаг является одним из наиболее важных для метода, так как он дает искомые результаты решения задачи анализа. Данный шаг не может быть формализован в общем виде. Аналитик должен понять связь и преобразование исходных данных в пространственную сцену и “откатить их назад”, чтобы сделать выводы об исходном объекте. Эффективность этого шага в значительной степени зависит от понимания аналитиком предметной области и его опыта работы в ней. В то же время этот шаг дает аналитику гораздо более глубокое и, по мнению Дж. Томаса, интуитивное (на уровне озарения) понимание данных и предметной области в целом.

Пример простого визуального анализа графического изображения показан на рис. 3. Нарисованный график анализируется и приводится ряд утверждений относительно непосредственно графика. Затем данные утверждения интерпретируются относительно исходных данных (из рис. 2).



Рис. 3. Простой пример визуального анализа

Поскольку взаимодействие между человеком и компьютером, является ключевым в решении этой задачи, метод визуализации становится интерактивным.

После того, как анализ завершен, результаты могут удовлетворить аналитика, тогда решение задачи анализа исходных данных завершено, или аналитик может решить вернуться к одному из предыдущих этапов решения задачи методом визуализации. Аналитик может решить вернуться к любому шагу любой из задач. Таким образом, метод становится итеративным.

Определяя природу метода визуализации, можно сказать, что этот метод является методом пространственного моделирования объекта рассмотрения. При решении задачи анализа данных с помощью метода визуализации объект рассмотрения моделируется в виде пространственной сцены, затем делаются утверждения о сцене, и данные утверждения интерпретируются по отношению к исходному объекту рассмотрения. Подводя итог, можно сказать, что метод представляет собой метод моделирования с пространственным моделированием объекта рассмотрения *Ob*. На основе объекта создается пространственная модель, а результаты анализа пространственной сцены интерпретируются по отношению к исходному объекту.

### 3. Визуальная аналитика в работах УНЛ «Научная визуализация» МИФИ

Целью создания лаборатории было использование подхода визуальной аналитики для содействия физическим кафедрам в их исследованиях. В результате этого сотру-

ничество был разработан и внедрен ряд пассивных и интерактивных приложений визуализации для анализа различных физических данных. Некоторые из них были разработаны в сотрудничестве с Британским национальным центром компьютерной анимации Борнмутского университета. Особое внимание привлекли к себе невидимые объекты — маленькие, как наноструктуры, или физически не видимые как скалярные или векторные поля. Разработанные приложения были статичными или анимированными, интерактивными или пассивными в зависимости от целей анализа и потребностей исследователей.

### 3.1. Применение в физических исследованиях

“Фокус” в физических исследованиях состоит в том, чтобы проиллюстрировать процессы таким образом, чтобы исследователь мог их увидеть. Как сказал Альберт Эйнштейн: «Если я не могу это изобразить, я не могу этого понять». Основная задача лаборатории заключалась в том, чтобы сделать невидимые вещи наглядными и легкими для восприятия.

Визуализация наноструктур является примером такой трансформации. Превращение чисел и формул в анимированную картинку является одним из лучших способов помочь ученым понять основные процессы в изучаемых объектах. Предоставление возможности изменять исходные данные (например, условия или количество исходных объектов) дает аналитику возможность решать свои задачи, не показывая числовые результаты расчетов и не тратя время на понимание того, что пошло не так. Вся необходимая информация и процессы наглядны, а вычисления выполняются в фоновом режиме.

Столкновение двух фуллеренов является хорошим примером такой задачи. Созданное программное обеспечение имело возможность изменения углов и скорости столкновения фуллеренов. Аналитик может изменять параметры, чтобы получить изображение, которое он хочет получить. На рис. 4 при ведено два видео изображения, на одном из них происходит соединение фуллеренов, а на другом они разлетаются. [9]

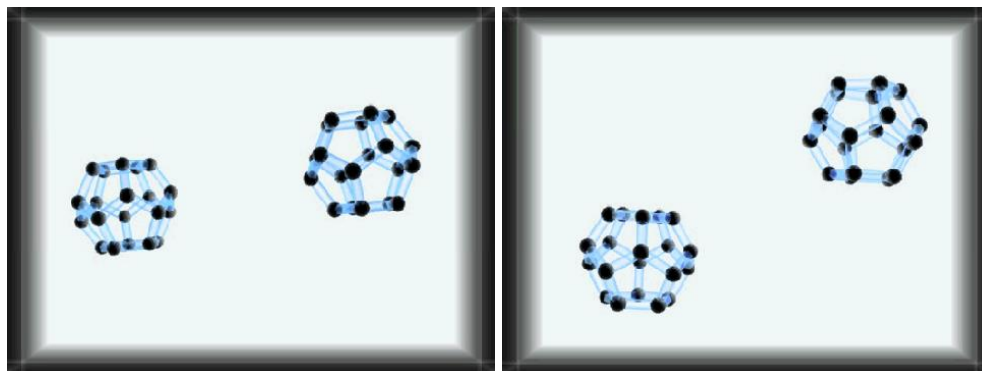


Рис. 4. Анализ столкновения фуллеренов.

Другим чрезвычайно важным объектом для визуализации являются различные поля. Наша лаборатория в сотрудничестве с физическими кафедрами работала как со скалярными, так и с векторными (тензорными) полями. Во всех случаях основной вопрос касался регионов с низкими или высокими значениями, а также нахождение эквивалентных зон. В этих исследованиях мы использовали как формы, так и оптические атрибуты, чтобы показать значения и направления полей.

Анализ поля параметров сверхпроводника является хорошим примером подобной визуализации. Сверхпроводник моделируется в соответствии с теорией Гинзбурга-Ландау. Разработанное программное обеспечение показывает текущие потоки, а цвет зависит от значений поля. Это приложение позволило физикам увидеть и понять направления и значения поля

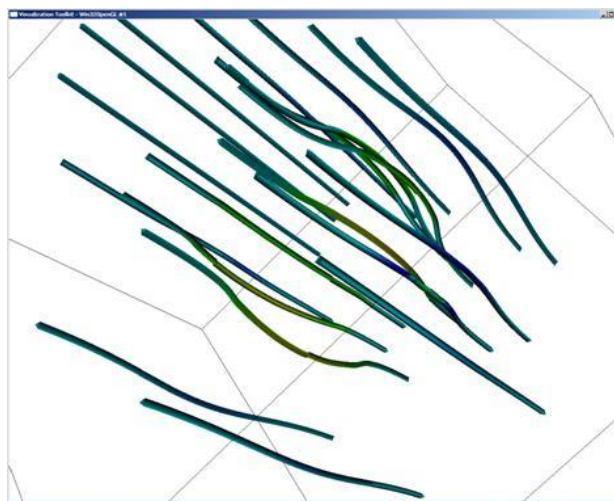


Рис. 5. Графическое представление линий тока векторного поля параметра порядка сверхпроводника второго рода

В некоторых случаях было необходимо сочетание двух объектов: нанобъекта и его поля. В этих случаях проблема заключалась в том, чтобы визуализировать оба компонента таким образом, чтобы они не блокировали друг друга, и можно визуально анализировать оба компонента одновременно.

Одним из проектов лаборатории была визуализация нанобъекта и его поля электронной плотности. Наше приложение сначала визуализировало нанобъект (в виде сфер и соединений), а затем добавило полупрозрачные изоповерхности, чтобы показать поле. Цвет поверхности зависит от значения поля. Этот вид визуализации позволяет исследователю видеть, как структуру, так и плотность поля вокруг него.

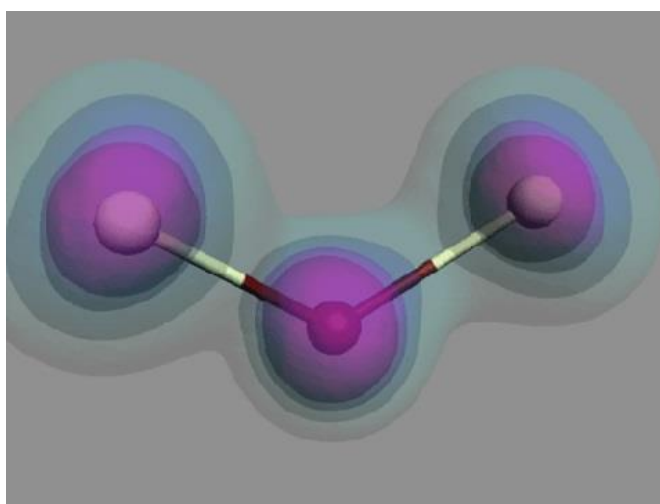


Рис. 6. Визуализация наноструктуры  $Cl_2O$  и ее поля электронной плотности.

### 3.2. Применение в финансах

Наша лаборатория принимала участие в проекте по анализу данных кредитных организаций, осуществляемом Федеральной службой по финансовому мониторингу Российской Федерации. Основной целью было найти кредитные организации, которые ведут себя подозрительно, и передать информацию для ручной и более глубокой проверки. Формализуя цель, можно сказать, что это была задача обнаружения аномалий. Из всех объектов рассмотрения аналитика интересуют объекты не похожие на остальные. [10] [11]

Для решения этой задачи была создана программа для визуального анализа многомерных данных. Основной идеей было создание дополнительных построений в многомерном пространстве. Если расстояние между двумя объектами в многомерном пространстве меньше, чем заданная аналитиком мера различия  $d$  ( $\rho(x_i, x_j) \leq d$ ), то мы строим отрезок в многомерном пространстве. Следующим шагом является проекция на трехмерное пространство и отображение объектов в виде сфер, а отрезков в виде цилиндров. Мы используем цвета цилиндров, чтобы показать расстояние между многомерными объектами.

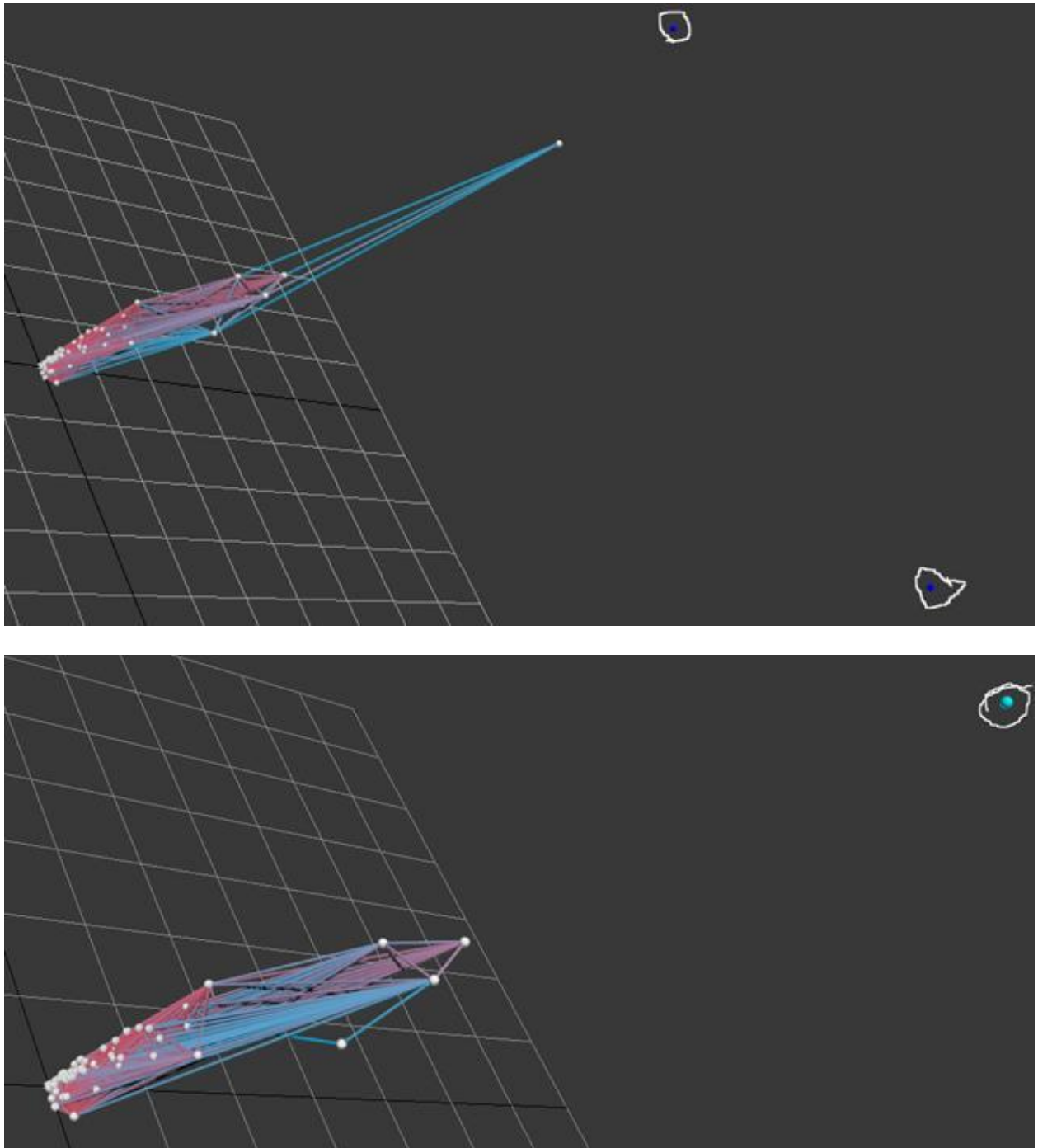


Рис. 7. Обнаружение аномалий с использованием метода визуализации.

### 3.3. Применение в мониторинге вычислительных

Вычислительные центры генерируют большой объем многомерных метаданных о своей работе. Основная проблема заключается в анализе этих данных, так как большая часть сообщает о нормальной работе, а интересующая часть скрыта под огромным количеством бесполезных данных. Наша лаборатория работала с группой из ЦЕРНа над этим проектом. Основной целью было выяснить, какие вычислительные центры замедляют работу и в каких центрах наблюдаются проблемы. [12]

Для этого было разработано два приложения - одно берет метаданные центра обработки данных, группирует данные с использованием некоторых автономных алгоритмов кластеризации и проецирует данные в трехмерное пространство. Кластеры в данной программе показаны с использованием цветов.

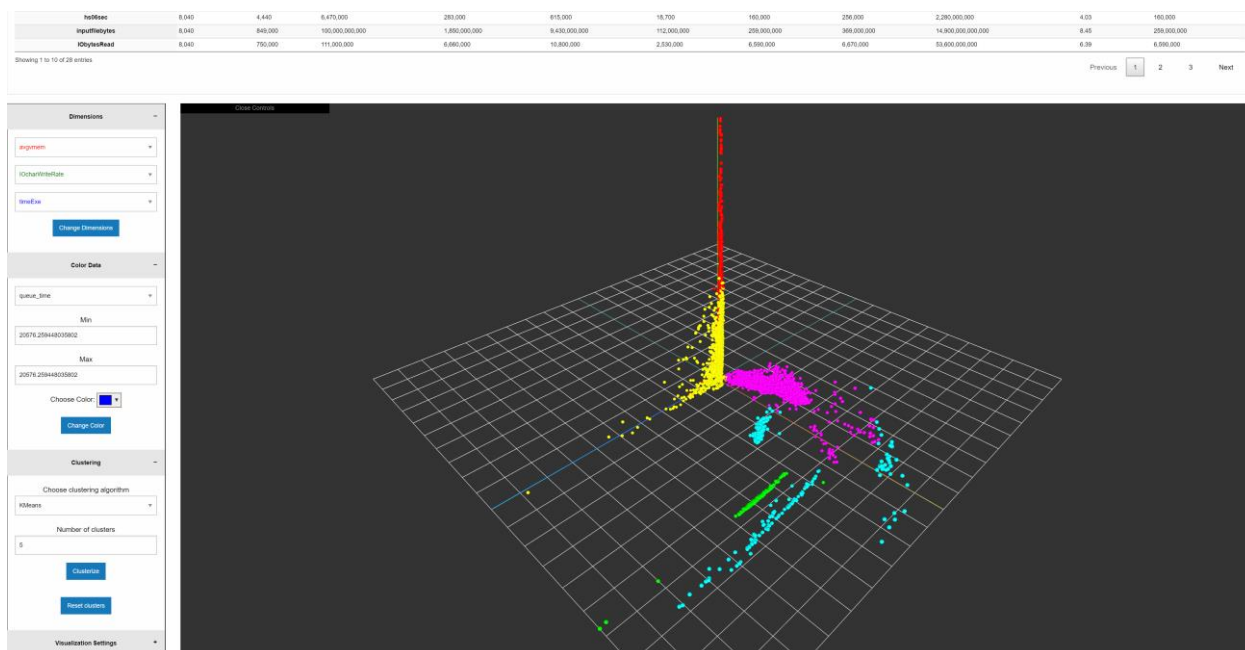


Рис. 8. Кластеризованные данные (более 8000 объектов с 28 параметрами, сгруппированными в пять кластеров с использованием алгоритма К-средних)

Второе приложение работает с сетевыми данными. Оно разработано для выявления проблем в связи между центрами обработки данных. Чтобы решить эту задачу, центры обработки данных были размещены на двух осях, и данные были визуализированы как сетка данных. С такой визуализацией аналитик ищет пики на сетке. Пример показан на рис. 9. Пиками являются красные и желтые точки.



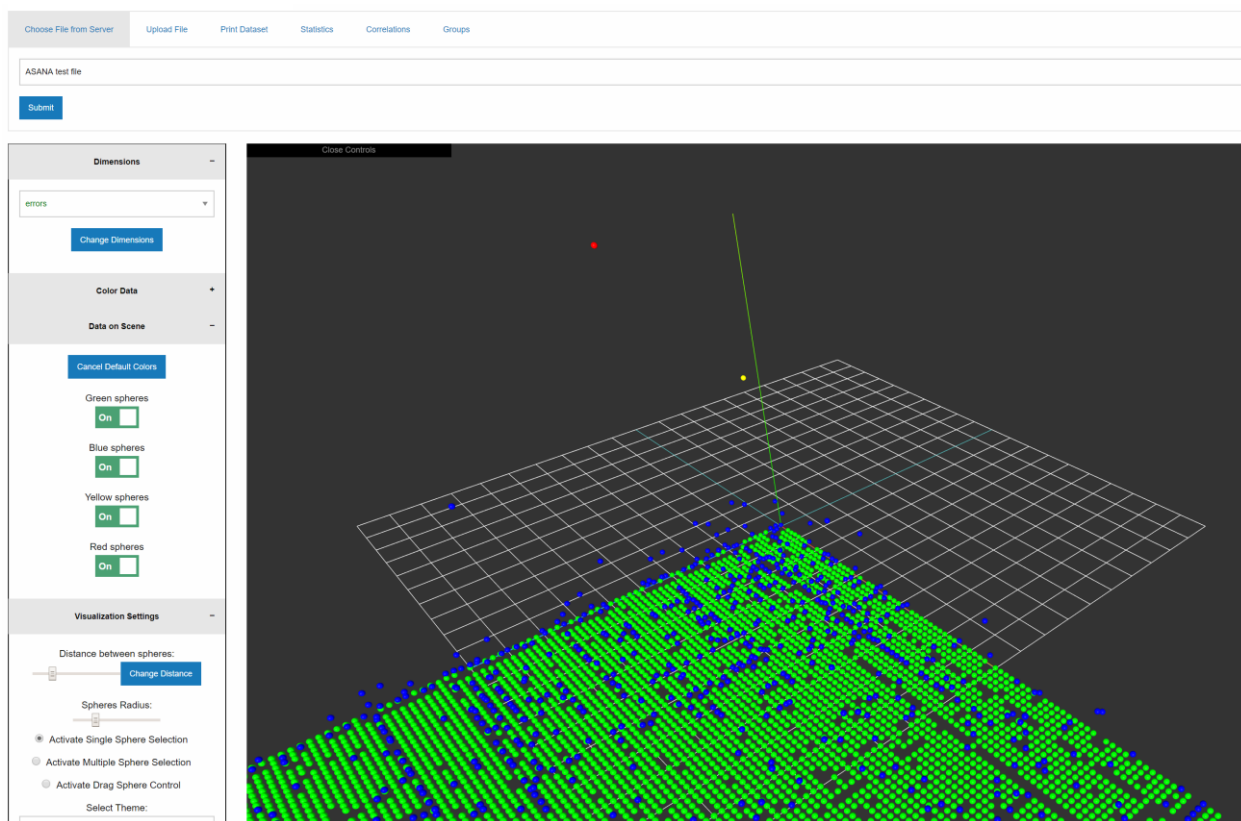


Рис. 9. Сетка данных с шипами.

### 3.4. Иные виды деятельности

Наша лаборатория проводит ряд мероприятий помимо исследований, но связанных с визуальным анализом. Первое — это обучение визуальному анализу и визуализации. В настоящее время на базе лаборатории преподаются два курса: базовый (визуальная аналитика) и углубленный (научная визуализация). Первый предназначен для магистрантов, а продвинутый — для аспирантов. [6]

Вторым видом деятельности является выпуск журнала «Научная Визуализация». Это электронный журнал открытого доступа с редколлекцией со всего мира. В настоящее время выходит пять номеров в год в сотрудничестве с двумя конференциями. Каждый выпуск состоит из десяти-двенадцати статей. Журнал индексируется в ряде баз данных: SCOPUS, RSCI и Compendex. Визуальный анализ является главной темой журнала: все статьи так или иначе связаны либо с визуальным анализом, либо с компьютерной графикой. [13]

### Заключение

Последние годы деятельность нашей лаборатории в основном сосредоточена на теории и разработке приложений для визуального анализа многомерных данных. [14] Проводимые исследования могут как использовать метод визуализации, так и быть не связанными с ним, но они всегда ориентированы на сочетание вычислений с интерактивной визуализацией. Используя эту технику, аналитик получает лучшее из обоих миров: может использовать совместно свое формально-логическое мышление, а также пространственно-образное мышление для решения сложных проблем.

## Список литературы

- [1] J. Thomas, K. Cook, V. Crow, B. Hetzler, R. May, D. McQuerry, R. McVeety, N. Miller, G. Nakamura, L. Nowell and P. Whitney, "Human—Computer Interaction with Global Information Spaces—Beyond Data Mining," *Digital Media: The Future*, pp. 32-46, 2000.
- [2] J. Thomas and K. Cook, *Illuminating the Path: Research and Development Agenda for Visual Analytics*, IEEE-Press, 2005.
- [3] M. L. Huang, J. Liang and Q. V. Nguyen, "A Visualization Approach for Frauds Detection in Financial Market," in *2009 13th International Conference Information Visualisation*, Barcelona, Spain, 2009.
- [4] A. A. Cárdenas, P. K. Manadhata and S. P. Rajan, "Big Data Analytics for Security," *IEEE Security & Privacy*, vol. 11, no. 6, pp. 74-76, 2013.
- [5] M. S. Khine, "Spatial Cognition: Key to STEM Success," *Visual-spatial Ability in STEM Education*, pp. 3-8, 2017.
- [6] V. Pilyugin, "Scientific Visualization Laboratory of NRNU MEPhI," NRNU MEPhI, [Online]. Available: <http://sv-journal.org/unl/>. [Accessed 25 11 2019].
- [7] V. Pilyugin, E. Malikova, A. Pasko and V. Adzhiev, "Scientific Visualization As Method Of Scientific Data Analysis," *Scientific Visualization*, vol. 4, no. 4, pp. 56-70, 2012.
- [8] D. Keim, F. Mansmann, J. Schneidewind, J. Thomas and H. Ziegler, "Visual Analytics: Scope and Challenges," *Visual Data Mining. Lecture Notes in Computer Science*, vol. 4404, 2008.
- [9] M. Strikhanov, N. Degtyarenko, V. Pilyugin, E. Malikova, M. Matveeva, V. Adzhiev and A. Pasko, "Computer Visualization Of Nanostructures Experience At NRNU "MEPHI"," *Scientific Visualization*, vol. 1, no. 1, pp. 1-18, 2009.
- [10] I. Milman, A. Pakhomov, V. Pilyugin, E. Pisarchik, A. Stepanov, Y. Beketnova, A. Denisenko and Y. Fomin, "Data Analysis Of Credit Organizations By Means Of Interactive Visual Analysis Of Multidimensional Data," *Scientific Visualization*, vol. 7, no. 1, pp. 45-64, 2015.
- [11] I. Milman and V. V. Pilyugin, "Interactive Visual Analysis of Multidimensional Geometric Data.," in *24 th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision WSCG 2016*, Plzen, Czech, 2016.
- [12] T. Galkin, M. Grigoryeva, A. Klimentov, T. Korchuganova, I. Milman, S. Padolski, V. Pilyugin, D. Popov and M. Titov, "The new approach to monitor the workflow management system ProdSys2/PanDA of the ATLAS experiment at LHC by using methods and techniques of visual analytics," *Scientific Visualization*, vol. 10, no. 1, pp. 77-88, 2018.
- [13] "Scientific Visualization," [Online]. Available: <http://sv-journal.org/>. [Accessed 25 11 2019].
- [14] O. Maslennikov, I. Milman, A. Safiulin, A. Bondarev, S. Nizametdinov and V. Pilyugin, "Development Of A System For Analyzing Of Multidimensional Data," *Scientific Visualization*, vol. 6, no. 4, pp. 30-49, 2014.

# Visual analytics and its use in the NRNU MEPhI “Scientific Visualization” laboratory activities

V.V. Pilyugin<sup>1,A</sup>, I. Milman<sup>2,B</sup>

<sup>A</sup> National Research Nuclear University MEPhI, Moscow, Russian Federation

<sup>B</sup> Individual researcher, USA

<sup>1</sup> ORCID: 0000-0001-8648-1690, [VVPilyugin@mephi.ru](mailto:VVPilyugin@mephi.ru)

<sup>2</sup> ORCID: 0000-0001-9705-9401, [Igal.Milman@gmail.com](mailto:Igal.Milman@gmail.com)

## **Abstract**

The article discusses visual analytics, which, according to J. Thomas, is understood as the solution of problems of data analysis using a facilitating interactive visual interface. Today, visual analytics is widely used in various fields – in research, design, management and others, due to the well-known predisposition of people to spatial thinking. Despite the wide usage, according to the authors, the theoretical aspects of visual analytics are not well developed at present, which is certainly a limiting factor in the development of visual analytics tools and its effective use in practice. It is declared in the article, that one of the most common of the forms of visual analytics is solution of data analysis problems using the visualization method. The visualization method is described in details and a number of works done by the “Scientific Visualization” laboratory NRNU MEPhI in this area are presented as an example of usage of the visualization method. The works were performed in different subject areas from physics to finance and monitoring.

**Keywords:** Visual analytics, data analysis, visualization method, modeling.

## **References**

- [1] J. Thomas, K. Cook, V. Crow, B. Hetzler, R. May, D. McQuerry, R. McVeety, N. Miller, G. Nakamura, L. Nowell and P. Whitney, "Human—Computer Interaction with Global Information Spaces—Beyond Data Mining," *Digital Media: The Future*, pp. 32-46, 2000.
- [2] J. Thomas and K. Cook, *Illuminating the Path: Research and Development Agenda for Visual Analytics*, IEEE-Press, 2005.
- [3] M. L. Huang, J. Liang and Q. V. Nguyen, "A Visualization Approach for Frauds Detection in Financial Market," in *2009 13th International Conference Information Visualisation*, Barcelona, Spain, 2009.
- [4] A. A. Cárdenas, P. K. Manadhata and S. P. Rajan, "Big Data Analytics for Security," *IEEE Security & Privacy*, vol. 11, no. 6, pp. 74-76, 2013.
- [5] M. S. Khine, "Spatial Cognition: Key to STEM Success," *Visual-spatial Ability in STEM Education*, pp. 3-8, 2017.
- [6] V. Pilyugin, "Scientific Visualization Laboratory of NRNU MEPhI," NRNU MEPhI, [Online]. Available: <http://sv-journal.org/unl/>. [Accessed 25 11 2019].
- [7] V. Pilyugin, E. Malikova, A. Pasko and V. Adzhiev, "Scientific Visualization As Method Of Scientific Data Analysis," *Scientific Visualization*, vol. 4, no. 4, pp. 56-70, 2012.
- [8] D. Keim, F. Mansmann, J. Schneidewind, J. Thomas and H. Ziegler, "Visual Analytics: Scope and Challenges," *Visual Data Mining. Lecture Notes in Computer Science*, vol. 4404, 2008.

- [9] M. Strikhanov, N. Degtyarenko, V. Pilyugin, E. Malikova, M. Matveeva, V. Adzhiev and A. Pasko, "Computer Visualization Of Nanostructures Experience At NRNU "MEPHI"," *Scientific Visualization*, vol. 1, no. 1, pp. 1-18, 2009.
- [10] I. Milman, A. Pakhomov, V. Pilyugin, E. Pisarchik, A. Stepanov, Y. Beketnova, A. Denisenko and Y. Fomin, "Data Analysis Of Credit Organizations By Means Of Interactive Visual Analysis Of Multidimensional Data," *Scientific Visualization*, vol. 7, no. 1, pp. 45-64, 2015.
- [11] I. Milman and V. V. Pilyugin, "Interactive Visual Analysis of Multidimensional Geometric Data.," in *24 th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision WSCG 2016*, Plzen, Czech, 2016.
- [12] T. Galkin, M. Grigoryeva, A. Klimentov, T. Korchuganova, I. Milman, S. Padolski, V. Pilyugin, D. Popov and M. Titov, "The new approach to monitor the workflow management system ProdSys2/PanDA of the ATLAS experiment at LHC by using methods and techniques of visual analytics," *Scientific Visualization*, vol. 10, no. 1, pp. 77-88, 2018.
- [13] "Scientific Visualization," [Online]. Available: <http://sv-journal.org/>. [Accessed 25 11 2019].
- [14] O. Maslennikov, I. Milman, A. Safiulin, A. Bondarev, S. Nizametdinov and V. Pilyugin, "Development Of A System For Analyzing Of Multidimensional Data," *Scientific Visualization*, vol. 6, no. 4, pp. 30-49, 2014.