# Using a Game Engine to Integrate Experimental, Field, and Simulation Data for Science Education: You Are the Scientist!

Jean-Yves Hervé, Brian Mullen, Tomás Francisco Yago Vicente, Christopher T.S. Allen, Charles Morace, and Ivan Otterness

Abstract—The purpose of this project is to use a game engine to integrate geo-referenced research data, whether experimental or simulated, to present it interactively to the user. Geo-referenced means that every image, video, or sound file, every pressure map, and every simulated temperature chart is attached to a specific point on a map or body. These data may also be timereferenced, so that different data sets may be available at the same location for different times of the day or seasons of the year. Target users for the interactive applications are high-school and college students who can then conduct their own "experiments" or "explorations" as a way to get exposed to the problems and methodologies of science and research. We use two examples of projects to illustrate the approach.

Index Terms—Science education, data integration, georeferenced data, game engine, interactive media.

## I. INTRODUCTION

SING games and play as means to teach "serious" topics to children and teenagers is not a new idea: A number of seminal works in the 1980's argued at length for the benefits of a ludic approach to teaching, in particular inwhat was not yet calledSTEM (Science, Technology, Engineering, Mathematics) fields ([1], [2]). Since 2000 there has been an increased awareness of games as a teaching tool ([3]-[5]). In particular, the vast potential of casual gaming for education purposes has been the focus of numerous publications and studies. One only needs to look at the figures announced by the gaming industry for the casual games market [6] to understand that the familiar gameplay mechanics combined with a simple interface can be used in teaching trigonometry, chemistry, or basic mechanics [7], [8]. It sometimes even seems as if gamebased education were the only way to "save" mathematics and science education [9]. A great deal of the work in that direction has taken the form of casual games that focus on teaching about the mechanics of a particular concept. In their simplest form such games are little more than interactive quizzes in which the player is asked to map a function to its derivative or fill in definitions [10]. In the more sophisticated series of educational games presented by [11], students need to apply simple results from calculus to solve problems such

as computing the area of a butterfly's wing or designing a roller coaster ride.

Although our group is also engaged in casual gaming projects for STEM and health education ([12], [13]), the focus of this paper is on ways to introduce teenagers to notions about the scientific process: What is it that a researcher does? What kind of data do scientists deal with? How do they interpret them, and how does this lead to seeking new data? To illustrate this article, we have been developing two sample applications and a general utility tool that showcase the key elements of our approach. Although both applications target very specific research problems, one from environmental science, one from medical science, the concepts presented here are applicable to other research problems and domains as well. The first application deals with a simplified problem of underwater exploration. The second application is concerned with the problem of knee arthritis and its early detection through the study of changes in the local geometry of the surface of knee articular cartilage. Both application domains are characterizedas is much of today's researchby the wide range of types and vast amount of data available to the researcher. In the case of our underwater exploration example, such data could include temperature and current measurements at different points in the ocean, possibly over a period of time, sonar sensing of different points of interest, a reconstruction of the ocean floor, video and audio recordings taken by underwater vehicles or divers. In the case of the knee arthritis application, anonymized data include classical X-ray images or 3D imagery obtained through an MRI scan (Figure 1) with clinical data about a patient's health history.

One important aspect of these data is that they are all georeferenced and time-referenced. A particular video is recorded atand corresponds to a specific location and time, Even if the "geo-tag" and "time-tag" can be ignored to exploit the generality of some of the data (e.g. a video of a particular type of fish could typically have been shot at other locations and at other times of the day or year, with similar results), some of that data is tightly connected to the time and place where it was collected, often through synchronous information gathered by other sensor modalities. For the knee arthritis problem, MRI and X-ray data are captured at a specific date, enabling the evolution of disease to be followed in a time sequence.

Our demonstration applications combine all the data available and supplement them with on-demand simulation data. When real data is not available, the application provides

J-Y Hervé, B. Mullen, T.F. Yago Vicente, C.T.S. Allen, and C. Morace are with the 3D Group for Interactive Visualization, Department of Computer Science and Statistics, University of Rhode Island, Kingston, RI 02881, USA. e-mail: {jyh,brian,tyago,callen,cmorace}@cs.uri.edu.

I. Otterness is with the School of Pharmacy, University of Rhode Island, Kingston, RI 02881, USA. e-mail: iotter@mail.uri.edu

Manuscript received June 27, 2010; revised July 07, 2010.



Fig. 1. Examples of knee MRI and X-ray data.

aspiring researchers an integrated system through which they can conduct their own research experimentation. With the underwater exploration, the aspiring researcher can plan an exploration mission, direct sensors to collect data, and analyze these data. With the knee arthritis application, it will be possible to request MRI scans of a patient's knee, compare such scans over a multiple-year period, and from the evolution of the cartilage's surface infer the patient's arthritis status.

Our main development environment is the Unity3D game engine [14]. We selected Unity because it combines support for a wide range of multimedia data formats, a very complete API (Application Programming Interface) accessible either in Javascript or C#, and a plugin architecture that allows us to add complex scientific computing modules and exploit in a Unity application the power of third party libraries such as QuickTime or the Mathematica kernel. Additionally, Unity is fairly accessible to artists, which greatly facilitates project development in a multidisciplinary group such as ours.

## **II. UNDERWATER EXPLORATION**

The premise of this application is that several data collecting missions involving surface and underwater crafts or teleoperated sensing devices have brought back large amounts of geo- and time-referenced data about a specific underwater site. In particular, a complete 3D reconstruction of the ocean floor at this location has been computed by combination of sonar, stereographic, and other sensory information. Detailed information about particular points of interest is provided under the form of video or audio recordings, higher-density local sonar or stereo-graphic scans, temperature and current measurements recorded by a sensor grid, etc.

Our aspiring researcher is placed in charge of a simulated underwater rover that can travel on the ocean floor and record data to be analyzed during and after the mission. Some of the data collected by the sensors during the mission may be accessible "live" to the researcher, with some transmission delays, but much of that data requires a great deal of processing before it can be visualized and analyzed, and will therefore only be accessible to the researcher after the mission has been completed. Other data may be potentially available but limits on the simulated rover's battery life and communication bandwidth will limit how much data the researcher will have access to during the mission. It is therefore critical to plan the mission carefully, taking into account constraints such as the range of operation of the rover and of its sensors and communication equipment. Problems with the initial plan may be detected from the live data available at the control station and selected to be viewed by the researcher, in which case the researcher may elect to make local adjustments to the plan to salvage the mission.

The rover's path can be planned on an overview map showing only information known prior to the exploration mission, after which the rover can be sent on its path. The researcher can choose what information to display in the different panels of the application. For example, in Figure 2, the main panel shows the estimated location of the rover on its path while smaller panels display information provided by the embarked sensors, some real, some simulated.



Fig. 2. The main panel shows the overview map of the site to allow the junior researcher to plan a path of exploration and data collection.

All of the small panels in Figure 2 show video streams, except for one panel that shows a "waterfall" display of multiple-beam sonar data.



Fig. 3. As the underwater rover travels at the bottom of the ocean, the application displays asynthetic "first person" view of the exploration site.

Information about the 3D profile of the ocean floor, local

geology, and underwater flora and fauna, allow the game engine to generate several "first person" views of the ocean floor, as it would be seen by cameras mounted on the rover. Figure 3 gives an example of simulated video presented in real time to the researcher as the rover follows its path at the bottom of the ocean. The rover's local orientation, and therefore the directions its on-board cameras and other sensors point to, are determined by the local geometry of the ocean floor. If unexpected obstacles appear in front of the rover in such a view, or if the researcher observes an element in the scene deemed worthy of further investigation, then control of the rover can be switched from automatic, path following mode to direct remote control mode (with all the problems inherent to this mode, communication latency in particular).

On-board sensors can be pointed to a specific point of interest and the data collected by these sensors can be visualized in one of the application's panels. Simulated sensor data accurately reflect the conditions at the current location of the rover and for the current time. For example, the simulated sonar correctly reproduces occlusion by elements in the scene. Consequently, if the researcher does not make a good exploration plan, and fails to correct the plan during the mission, the data collected might prove invalid, or insufficient for a proper post-mission analysis.

When real data that is appropriate to the current location and time is available, and has been selected by the researcher to be "transmitted" live to the control station, that data is displayed instead of synthetic data. For example, Figure 4 shows in the main panel a video recorded by a diver at the current location and now presented as coming from one of the rover's side cameras.



Fig. 4. The application displays in the main panel real video footage available for that location and time.

After the exploration mission has been completed, the data collected by the different sensors selected by the researcher is now made available for analysis and visualization. We discuss in Section 4 this aspect of research and how we use a game engine for that purpose but, before we leave our underwater explorer to move on to another application, we can see another important feature of our application: performance evaluation.

Just as the game engine was recording sensor data, whether real or simulated, during the the rover's mission, all the actions of the researcher in the planning and execution of the mission are recorded as well, and can be viewed for post-mission evaluation. In that mode, additional views such as the "third person" view shown in Figure 5 become available. This mode allows the researcher to review mistakes made in the planning of rover's path or sensing schedule, and how adjustments should have been made based on available information.



Fig. 5. A "third-person" view of the exploration rover is also available for post-mission evaluation.

### **III. EXPLORING KNEE OSTEOARTHRITIS**

We are developing an application that will allow an aspiring researcher to access longitudinal X-ray and MRI data (typically recorded 1 year apart). Some of the individuals are healthy, some are already suffering from advanced osteoarthritis, some are just beginning to show signs of the disease but have not necessarily reached pain levels that would trigger a visit to a physician. Clinical data about these patients giving gender, age, height, weight, knee pain score, and other relevant parameters are linked to the images. Using the MRI images, they can explore the knee anatomy in detailexamining meniscus, ligaments, boney and cartilage surfaces, learning the anatomy and selecting measurables that they can test for linkage with either the presence of disease or its evolution. Then they can explore in more depth the cartilage and boney surfaces and the space filling aspects of the cartilage itself. (Meniscal, ligament and tendon modules can be added later).

MRI data has been "segmented" to identify the boundaries of the inner surface (bone interface) and outer surface of knee articular cartilage, as shown in Figure 6.

Using the segmented cartilage data we compute local geometric characteristics such as the thickness of the cartilage or the curvature at a point of the inner or outer surface. This information can then be rendered visually, as in Figures 7 and 8, which show the a patient's inner surface (bone interface) of the lateral and medial tibia rendered as a colorless wireframe mesh and then colored based on local thickness.



Fig. 6. Segmentation of the cartilage's inner and outer surfaces in the MRI data.



Fig. 7. Inner surface of the medial tibia cartilage (wireframe rendering).

It is intuitively obvious that osteoarthritis, which can result in actual damage of the cartilage, will affect eventually the local geometry of the cartilage. One of the hypotheses that guided the research on which this interactive application is based is that some changes in the local geometry are an early indicator of osteoarthritis. Osteoarthritis, however, is not the only thing that can change cartilage geometry: age, significant weight increases, load bearing exercise, trauma, etc., can also have observable effects. Our aspiring researcher will learn to differentiate between normal and abnormal changes observed between annual measurements of the same patient.

Besides local geometrical characteristics of the cartilage's



Fig. 8. Surface of the lateral and medial tibia cartilage colored according to local thickness.

surface, the application also "registers," that is puts in correspondence, measurements made for the same patient over a multiple-months (typically, one year), interval. This registration process allows then to compare local changes of the characteristics computed: thickness, Gaussian and maximum curvatures, etc. Furthermore, it also allows to "predict" future changes in cartilage shape by extrapolating changes observed in the last measurements, and thus generate visualizable mesh data, and even possibly MRI data when real data is not available. Existing MRI data can also be used to produce "new" MRI data obtained by resampling the initial MRI image stack along a slightly rotated grid (reflecting a small difference of orientation of the patient's knee in the MRI scanner).

The aspiring researcher has access to all the data, whether real or simulated, of a patient, and to all the results made possible by the registration of 3D data over time. For example, Figure 9 shows a simulated image of changes in local cartilage thickness observed for a patient over a one-year period.



Fig. 9. Simulated view of changes in cartilage thickness over a one-year period.

This rendering can be rotated to better reveal regions where the most significant changes occur (Figure 10). The application also provides global information (such as average change over particular regions of the knee cartilage).



Fig. 10. Simulated view of changes in cartilage thickness over a one-year period.

The aspiring researcher can study the evolution of the patient's cartilage over the last two or three years, compare with the data reference patients who are known to have knee osteoarthritis and with the data of reference healthy patients, in order to develop their own diagnostic

## IV. A TOOL FOR DATA ANALYSIS AND VISUAL DATA MINING

A game engine such as Unity also proves to be an invaluable tool in developing applications that address another important aspect of the research experience: the analysis and visualization of the data collected by our underwater explorer or knee arthritis clinician. Classical statistics has developed a number of sophisticated tools for analyzing data, but these tools typically suffer from imposing very strict conditions in order to be applicable. For example, because a great deal of multivariate calculus relies on linear algebra computations, a common constraint is that of linear relations between the data, (or at least of a known nonlinear relation that can be "inverted" to produce linearly related data). Our data can typically not be analyzed directly by such tools because there are often non-modeled nonlinear relations between the different types of data (in fact, such relations are precisely what we want to discover) and the data sets are too large to be easily clustered by automatic techniques. Our objective in this section is to explore an approach for clustering large data sets and determining nonlinear relations between data so that the data can be pre-processed and fed to classical statistical tools. This approach is based on the interactive visualization and direct manipulation of large data sets. Information visualization deals with data sets lacking inherent 2D or 3D semantics and therefore also lacking a standard mapping of abstract data onto the physical space of the paper or screen. A number of wellknown techniques visualize (partially) such data sets, including x-y plots, line plots, and histograms. These techniques are

useful for data exploration but are limited to relatively small low-dimensional data sets. A large number of information visualization techniques have been developed over the past decades, allowing visualizations of ever larger and more complex, or multidimensional, data sets ([15]-[17]). Among these, geometrically-transformed display techniques aim at finding useful information by using geometric transformations and projections. There is, however an infinite number of possibilities to project high-dimensional data onto the two dimensions of a standard display. "Projection pursuit" [18] attempts to locate projections that satisfy some computable quality of interest. A particular projection pursuit technique known as the GrandTour [19] aims at automatically finding interesting projections or at least helping the user to find conclusion by exploiting the remarkable capabilities of the human visual system: Users can scan, recognize and recall images rapidly, and can detect subtle changes in size, color, shape, movement, or texture. They can extract efficiently "group" information: alignment, symmetry, "sameness," and "togetherness." Most geometrically-transformed display techniques fail to exploit human perception because they present static representations. 3D-to-2D projections typically exhibit ambiguities that can only be resolved by a change of point of view, which is what the GrandTour proposes. The effectiveness of this technique is limited by the fact that viewpoint changes are passive, that is, controlled by the system rather than by the user. However, it has been clearly established by Bajcsy [20] and others that active vision is an essential aspect of scene understanding. It is by actively changing their viewpoint, eye vergence, focus, etc. that viewers acquire a better understanding of what they are looking at (even when what they are looking at is a flat picture or graph). We use interactive visualization as a tool to present the relationship between a discrete (binary) response and a set of explanatory variables. The visualization system we present allows users to "manipulate" directly, dynamically, and interactively their data set. The researcher assigns a visual role to different "columns" of a data set, and data points are displayed as colored objects in a user-controllable panel. Obvious visual roles are the coordinates and the color of the graphical object representing the data point. We have also explored the use of orientation, glyphs, and local animation as a way to convey additional information but at the time of this writing, evidence that this additional information facilitates analysis is inconclusive.

This new system opens up interesting possibilities for exploring data visually, through its better exploitation of the human visual system. The ability of humans to detect subtle changes in size, color, shape, movement, or texture, and, beyond that, to extract "group" information (alignment, symmetry, "sameness") are considerably superior to that of any of today's computer-based systems. This remarkable ability can be somewhat harnessed by letting the user's visual system guide the choice of transformations applied to the data set. The software tool we present here adds this new dimension to the visualization of large and complex data sets by allowing users to manipulate directly and interactively the data. In this respect, a significant added value of a game engine over a regular graphical library is the integration with the GUI and with a physics engine facilitating direct data manipulation. The user can dynamically change the visual role assigned to the different data of the data set, and quickly rotate the data points to make "interesting" alignments or groupings more visible. Figure 11 shows examples of assignments of visual roles to the elements of a data set, seen from different viewpoints.



Fig. 11. Examples of direct visualization of a data set using different visual roles and viewpoints.

Interactive, direct manipulation of the data set makes interesting groupings of data points "jump" at the user, when observed with the proper visual role assignment and viewpoint (see Figure 12). The data can then be clustered into such interesting groups of data points that can be visualized separately, with a different visual role assignment to reveal finer groupings. Naturally, even though we almost present here the assignment of a visual role as a blind, almost random search for a "lucky" combination, a good understanding of the data will lead to a much faster identification of suitable visual role assignments, would it be only through the non-consideration of trivial relations (such as the one between height and weight of a "typical" patient).

A group taking the appearance of a curved surface is indicative of a nonlinear relation between the data playing the visual role of X, Y, and Z coordinates. A simple fit to a parametric equation (for example, to a low-degree polynomial function) can be performed on the data points of the group to reduce the dimension from three to two, the corresponding columns in the data set being replaced by the new coordinates in the 2D parameter space of the fitted sheet. The transformation applied is recorded, to be later provided to the computational statistics module that will analyze the data for that particular cluster. In effect, the interactive visualization application is using gestalt perception abilities of the human visual to guide the operation of the statistical modules.

## V. CONCLUSION

In this article, we presented two applications that illustrate our approach to using video games and gaming technology to teach high school and college students about research and the research process. The first application simulates underwater exploration while the second application deals with the early detection of knee osteoarthritis. In both cases we use a game engine to integrate diverse geo- and time-referenced data (images, video and audio documents, measurement records), some of which was collected by real sensors or corresponds to



Fig. 12. A fortuitous combination of visual role assignment and viewpoint that reveals interesting groupings in the data set.

real patient data, and some of which is the result of simulation computations. The game engine allows us to offer interactive research challenges to students and aspiring researchers by exposing them to the research process. In particular we highlight both the collecting and the analysis of data. Additionally, the game engine allows students to find by visualization linear and non-linear relationships in the data. This naturally leads them into more complex analyses with the statistical package.

## ACKNOWLEDGMENT

Support for this work by by an NSF Research Experience for Undergraduates grant (award 0649059) and by a Partnership award from the University of Rhode Island is gratefully acknowledged.

#### REFERENCES

- S. Papert, Mindstorms: Children, Computers, and Powerful Ideas. New York, NY: Basic Books, 1981.
- [2] H. Ellington, Games and Simulations in Science Education. London, UK: Kogan Page Ltd, 1981.
- [3] Prensky, Digital game-based learning. New York, NY: McGraw-Hill, 2001.
- [4] J. Weld, *The Game of Science Education*. Boston, MA: Allyn and Bacon, 2003.
- [5] D. Oblinger, "The next generation of educational engagement," *Journal of Interactive Media in Education*, vol. 8, pp. 1–18, 2004.
- [6] J. 2007-Tams, Ed., Casual Games Market Report Casual Games Association, 2007. [On-Summary. line]. Available: http://www.casualconnect.org/newscontent/11-2007/CasualGamesMarketReport2007\_Summary.pdf
- [7] I. S. Group, "casual' video games play encouraged by parents, grandparents," 2007. [Online]. Available: http://www.infosolutionsgroup.com/pdfs/video\_games\_encouraged.pdf
- [8] C. Bisson and J. Luckner, "Computer games: Increase learning in an interactive multidisciplinary environment," *Journal of Educational Technology Systems*, vol. 24, no. 2, pp. 195–205, 1996.

- [9] D. Shaffer and J. Gee, "Before every child is left behind: How epistemic games can solve the coming crisis in education," Wisconsin Center for Education Research, University of Wisconsin-Madison, 2005.
- [10] "Calculus chapter 1." 2010. [Online]. Available: http://www.studystack.com/menu-265042
- [11] R. Kowalczyk and A. Hausknecht, "Use games to motivate your calculus students," in International Conference on Technology in Collegiate Mathematics, Atlanta, GA, Nov. 2000.
- [12] T. Yago Vicente, B. Mullen, T. Mather, and J.-Y. Hervé, "Casual gaming as a means to raise awareness of tick-borne disease risks," in 3rd Annual International Conference on Computer Games, Multimedia and Allied Technology, Singapore, Apr. 2010.
- [13] "3d group for interactive visualization, university of rhode island," 2010. [Online]. Available: http://www.3dgroup.cs.uri.edu "Unitv: Game development tool," 2010. [Online].
- "Unity: Game development tool," [14] Available: http://unity3d.com/
- [15] D. Keim, "Visual exploration of large data sets," Communications of the ACM, vol. 44, no. 8, pp. 176-188, 2001.
- [16] T. Soukup and I. Davidson, Visual Data Mining: Techniques and Tools for Data Visualization and Mining. Indianapolis, IN: John Wiley & Sons, 2002.
- [17] E. Tufte, The Visual Display of Quantitative Information, 2nd ed. Cheshire, CT: Graphics Press, 2001.
- [18] P. Huber, "Projection pursuit," The Annals of Statistics, vol. 13, no. 2, p. 435474, 1985.
- [19] D. Asimov, "The GrandTour: A tool for viewing multidimensional data," SIAM Journal on Scientific and Statistical Computing, vol. 6, pp. 128-143. 1985.
- [20] R. Bajcsy, "Active perception," vol. 76, no. 8, pp. 96-1005, 1988.



Jean-Yves Hervé is an Associate Professor at the Dept. of Computer Science and Statistics, University of Rhode Island. He is Ingénieur de l'École Centrale de Lyon, France (1985) and received his Ph.D. in Computer Science from the University of Maryland at College Park (1993), while doing his research at the Computer Vision Laboratory. From 1994 through 2000 he was an Assistant Professor of Computer Engineering at École Polytechnique de Montréal where most of his work dealt with telerobotics and

mining automation. Since joining URI his main research areas have been: simulation and visualization, computer vision, and bioinformatics. One important aspect of his work is the involvement of undergraduate students in research project.



Brian Mullen graduated from the University of Rhode Island with a BSc in Computer Science and a Minor in Digital Art. He also earned a certificate in Digital Media from The ICPNM Academy in Providence, RI. Brian has been Project Manager for the 3D Group for Interactive Visualization at the University of Rhode Island since 2005. His main areas of focus are scientific modeling & animation, web & print media, and digital content creation.



Tomás Francisco Yago Vicente received his Spanish Higher Engineering Degree in Computing from Universidad de Zaragoza. He is currently a Research Engineer at the 3D Group for Interactive Visualization, Department of Computer Science and Statistics, University of Rhode Island. His research interests include Visualization, Medical Imaging and Computer Vision.



Christopher T.S. Allen received his BS degree in Computer Science from the University of Rhode Island in 2009. He has been with the 3D Group for Interactive Multimedia since 2005 acting in various roles on a number of research projects. His research interests include computer vision, modeling and simulations, numerical analysis, and image processing.



Charles Morace is currently an undergraduate student at the University of Rhode Island double majoring in Computer Science and Applied Mathematics. He is an active member of the 3D Group for Interactive Visualization research group at URI and is interested in Computer Graphics and Computer Vision research.



Ivan Otterness received his B.A degree in Chemistry from St. Olaf College and his M.S. in Physical Chemistry and his Ph.D degree in Biophysics (1968) from the University of Southern California. He worked on the discovery of drugs to treat arthritis from 1971 to 2000 at Pfizer Inc. He is currently at the School of Pharmacy at the University of Rhode Island where he works on clinical methodology for testing arthritic drugs. In addition to his patents, he has published over 100

peer reviewed papers and edited 4 books. He has served on the editorial board of arthritis and inflammation journals, as President of the Inflammation Research Association, as a member of the US Pharmacopeia Committee of Revision and as a Director of the Canadian Arthritis Network.