Exploration and Analysis of Image-based Simulation Ensembles

M. Dahshan[1], T. L. Turton[2] and N. Polys[3]

[1] University of North Florida, USA
[2] Los Alamos National Laboratory, USA
[3] Virgina Tech, USA

Abstract

Scientists run simulation ensembles to study the behavior of a phenomenon using varying initial conditions or input parameters. However, the I/O bottlenecks hinder performing large-scale multidimensional simulations. In situ visualization approaches address the variability of I/O performance by processing output data during simulation time and saving predetermined visualizations in image databases. This poster proposes a visual analytics approach to exploring and analyzing image-based simulation ensembles, taking advantage of semantic interaction, feature extraction, and deep learning techniques. Our approach uses deep learning and local feature techniques to learn image features and pass them along with the input parameters to the visualization pipeline for in-depth exploration and analysis of parameter and ensemble spaces simultaneously.

CCS Concepts

• ***Computing methodologies*** $\rightarrow$ *Scientific visualization; Visual analytics; Image processing;*

1. Introduction

The rapid development of supercomputers empowers scientists from many scientific fields to run ensembles that could go beyond exabyte scales. Scientists carry out an ensemble to investigate the behavior of simulated phenomena in different states, find out the commonalities and differences between ensemble members, and explore parameter sensitivity and optimization. However, I/O bottlenecks make it hard to store massive volumes of data generated by large-scale simulation ensembles posing challenges to ensemble exploration and analysis [KI20, RR20].

To guarantee a high-quality analysis for large-scale ensembles, migrating analysis from traditional post-processing techniques to in situ approaches is needed to harvest the difference between the computing power of supercomputers and the lack of I/O capability. Since the output of in situ approaches is usually orders of magnitude less than simulation ensemble raw output. Image-based approaches have emerged as a promising technique for in situ analysis and visualization. Cinema framework [AJO*14] is an image-based in-situ approach that captures, stores, and analyzes exascale simulation data. It stores data abstracts, such as data parameters and images, at different camera viewpoints, facilitating instantaneous access to many data views. Cinema helps in browsing, managing, and exploring simulation data. Although cinema has demonstrated success in different analysis scenarios, it has potential shortcomings in analyzing and exploring simulation ensemble parameter and ensemble spaces. This results in difficulties in understanding and exploring major trends and correlations in an image-based simulation ensemble.

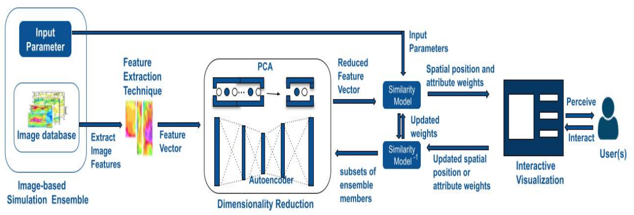

Figure 1: *The workflow has three main components: a) Input source: input parameters and image features b) Similarity models translate parametric and semantic interactions into manipulations of model parameters leading to a new visualization. c) Multi-linked visualizations.*

In this paper, we propose an interactive visual analytics approach that extends a current visualization tool, Graphically-Linked Ensemble Explore (GLEE) [DPJP20], extensively by incorporating the Cinema framework to create Cinema-GLEE (C-GLEE). This requires 1) added steps to extract significant features from image databases and combine them with simulation input parameters ; and 2) rephrasing GLEE's visualization pipeline to assure user interactions and model specifications are coherent. With C-GLEE, scientists gain a seamless, easy-to-use tool for analyzing, exploring, comparing, and browsing image-based ensembles interactively. In turn, scientists have opportunities to identify visually complex insights, such as correlations among the ensemble members, optimal parameter settings, and parameter sensitivities.

DOI: 10.2312/evp.20221128

2. C-GLEE

Our approach starts with an ensemble E $=\{e1,...,e_N\}$ of N members. Each member e_i $(i \in \{1,...,N\})$ is a collection of input parameters and an image-database representing simulation output from K camera positions; $e_i = \{inputs, I_1, ..., I_K\}$. Figure 1 shows the workflow of the proposed approach and the different visual analytics operations supported.

2.1. Image-based Simulation Ensemble Preprocessing

Moving from numeric simulation ensembles to image-based simulation ensembles requires extracting information or "features" from the image database to provide a basis for the interactive visualization pipeline. Relying on a single feature extraction technique may not be sufficient to extract representative features of the image for analysis as extracted features may have noise or unessential features. For example, global extraction techniques are more prone to cluttering and occlusion [NA19]. In contrast, local extraction techniques generate a high dimensional feature descriptor vector that could incur high computational costs, noise, and redundant information [ZZS*17]. On the other hand, deep learning techniques do not fit in scientific domains due to data scarcity, making it hard to have enough labeled data for supervised models or training data for unsupervised models [DT18].

In this paper, we adopt a two-level feature extraction approach. Our approach starts by applying a local feature extraction technique to the image database extracting a list of feature descriptors representing each image. Then, a dimensionality reduction (DR) technique is used to extract essential features from feature descriptors, dropping non-essential features. The reduced feature descriptors and input parameters are then passed as input to C-GLEE's visualization pipeline. The challenge in choosing a DR technique lies not only in discovering the underlying manifold structure and minimizing information loss during data mapping but also in capturing the meaning of scientists' interactive intents when using Semantic Interaction(SI) for ensemble exploration and analysis. Therefore, we investigate the differences between a traditional DR technique like PCA and a deep learning model like autoencoder(AE) in reducing the feature vector produced by local feature extraction.

2.2. C-GLEE Visual Interface

C-GLEE is a multi-linked visual interface following the principle of "overview first, zoom and filter, then details-on-demand" [Shn03].
Ensemble View (Figure 2a): Ensemble members are spatially arranged in a 2D space using a weighted dimensional projection technique based on input parameters, reduced visual features, and an associated weight vector. Weight vector elements are assigned an initial weight of (1/ l), where l represents the number of input parameters and the reduced visual features. Ensemble members are projected such that members with similar weighted attributes are grouped near each other while dissimilar ones sit farther apart. Ensemble view supports multiple interactions: observation-level interaction (OLI), member selection, and zooming. OLI is an interaction based on principles of SI. It allows scientists to convey their intuition and expertise by directly manipulating subsets of members based on a hypothesized similarity between them. OLI

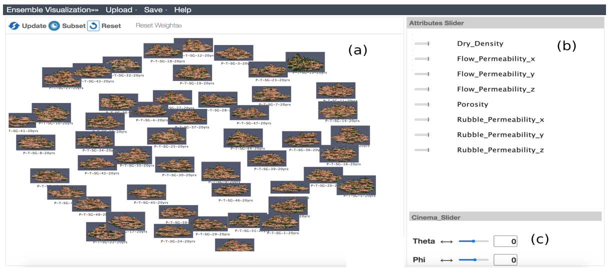

Figure 2: *C-GLEE's main interface: (a)The Ensemble view displays images representing ensemble members, spatially arranged in 2D space. (b)The Parameter view explores the influence of parameters on the ensemble by adjusting the weights to sliders. (c) The Cinema view is used to select a specific view angle from the Cinema image database.*

generates a new projection using a semi-supervised metric learning model that learns a new weight vector based on the similarities between manipulated members, updating ensemble view and weights on parameter view sliders. This helps scientists to find the commonalities and differences among ensemble members.

Parameter View (Figure 2b): Weights representing input parameters are displayed on a horizontal slider. The weights on the slider represent the importance of input parameters within the model. Scientists manipulate the slider values by performing a parametric level interaction (PLI). PLI results in an updated projection based on the updated weights, which allows scientists to understand and explore parameter sensitivity.
Cinema View (Figure 2c): The Cinema sliders are used to explore essential features in 3D ensembles in real-time. This helps in comparing and contrasting different ensemble members from different viewpoints, which creates an opportunity for new discoveries by exploring these viewpoints. Manipulating sliders results in resetting the weight vector and generating a new projection with new image features representing the new angle, unlike GLEE which only shows the image corresponding to the new angle without generating a new projection.
We evaluated our proposed approach with geoscience domain experts using two geological datasets (i.e., oilfield wastewater disposal [PMTC18] and CO2 sequestration [PFPM14].

3. Conclusion and Future work

This paper briefly introduces a visual analytics tool (C-GLEE) for exploring image-based simulation ensembles. Scientists' preliminary feedback showed that our tool led to the same conclusions derived from regular analysis methods in significantly less time. C-GLEE's interaction techniques find new insights and discoveries that are hard to find using traditional methods. Moreover, the reduced features generated by DL models produced visualizations that preserve complex intrinsic structures in the ensemble than the traditional DR techniques leading to more effective exploration. We plan to explore other Cinema database features in C-GLEE, such as time and non-image artifacts, for future work.

References

[AJO*14] AHRENS J., JOURDAIN S., O'LEARY P., PATCHETT J., ROGERS D. H., PETERSEN M.: An image-based approach to extreme scale in situ visualization and analysis. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis* (Piscataway, NJ, USA, 2014), SC '14, IEEE Press, pp. 424–434. 1

[DPJP20] DAHSHAN M., POLYS N. F., JAYNE R., POLLYEA R. M.: Making sense of scientific simulation ensembles with semantic interaction. In *Computer Graphics Forum* (2020), vol. 39, Wiley Online Library, pp. 325–343. 1

[DT18] DARA S., TUMMA P.: Feature extraction by using deep learning: A survey. In *2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)* (2018), IEEE, pp. 1795–1801. 2

[KI20] KAWAMURA T., IDOMURA Y.: Improvement in interactive remote in situ visualization using simd-aware function parser and asynchronous data i/o. *Journal of Visualization 23*, 4 (2020), 695–706. 1

[NA19] NIXON M., AGUADO A.: *Feature extraction and image processing for computer vision*. Academic press, 2019. 2

[PFPM14] POLLYEA R. M., FAIRLEY J. P., PODGORNEY R. K., MCLING T. L.: Physical constraints on geologic co2 sequestration in low-volume basalt formations. *Bulletin 126*, 3-4 (2014), 344–351. 2

[PMTC18] POLLYEA R. M., MOHAMMADI N., TAYLOR J. E., CHAPMAN M. C.: Geospatial analysis of oklahoma (usa) earthquakes (2011–2016): quantifying the limits of regional-scale earthquake mitigation measures. *Geology 46*, 3 (2018), 215–218. 2

[RR20] RIBÉS A., RAFFIN B.: The challenges of in situ analysis for multiple simulations. In *ISAV'20 In Situ Infrastructures for Enabling Extreme-Scale Analysis and Visualization*. 2020, pp. 32–37. 1

[Shn03] SHNEIDERMAN B.: The eyes have it: A task by data type taxonomy for information visualizations. In *The craft of information visualization*. Elsevier, 2003, pp. 364–371. 2

[ZZS*17] ZHEN X., ZHENG F., SHAO L., CAO X., XU D.: Supervised local descriptor learning for human action recognition. *IEEE Transactions on Multimedia 19*, 9 (2017), 2056–2065. 2