GPU-Assisted Visual Analysis and Categorization of Ensemble Conflict

Donald W Johnson*
Mississippi State University

T.J. Jankun-Kelly†
Mississippi State University.

Abstract

Analysis of multiple overlapping data scenes is a challenging problem with tension between clearly identifying and exploring significant overlaps & conflicts. Two areas where this problem occurs is when dealing with ensemble data from physical event simulation and when viewing multiple flood scenes that occur in an area of interest. In order to allow easier analysis of scenes with multiple overlapping data layers, we introduce a visualization system designed to aid in the analysis of such scenes. It allows the user to both see where different data sets agree, and categorize areas of disagreement based on participating surfaces in each area. The results are stable with regard to render order and GPU acceleration via OpenCL allows interaction with data large datasets. This interactivity is further enhanced by data streaming which allows datasets too large to be loaded directly onto the GPU to be processed. After demonstrating our approach on a diverse set of ensemble datasets, we provide feedback from expert users.

Index Terms: H.5.2 [User Interfaces]: Graphical user interfaces—[D.2.12]: Interoperability—Data mapping

1 Problem

Flooding is one of the most frequently occurring natural disasters; flood modeling is one tool used to mitigate their effects. These models can be used to predict how flood control projects would influence both historical and future floods; however, as the number of possible mitigation scenario's increases, the number of modeled outputs increases at an exponential rate. Moreover, in such disaster scenarios, disagreements or conflicts in the model ensembles are often more important than where they occur—a cautionary approach too flood planning is often warranted. Due to the multiplicity of model simulations, their integration and the identification of conflict is a challenge To address these difficulties, we propose a scalable visualization system that is designed to allow rapid analysis of multiple data surfaces. Although the motivation for designing this system was for flood scene analysis, it can be applied to any dataset where there are multiple overlapping data fields, such as those produced by weather model ensembles or other ensemble simulations of physical events. Our system quickly allows exploration of agreement and conflict in the overlapping surface with interactive exploration of the combinatorial space of possible interactions. Using an overview+detail approach, we also facilitate drilling down to the diverse collection of overlaps for any number of interactions. By using the GPU for computation and visualization, the systems increases the amount of user flow [1] by reducing mental state switches or iterative search enforced by standard techniques.

Current approaches to analyzing multiple overlapping spatial scalar datasets are costly in time and money. One contributing factor to this cost is that GIS tools display surfaces with ordered painting, where first one surface is drawn then the next and so forth until all surfaces have been displayed. This, however, causes surfaces that are displayed later to obscure surfaces that are earlier in the rendering sequence. Reordering surfaces and transparency are ways that a user can attempt to view obscured information, but such reordering becomes cognitively prohibitive as the number of surfaces/ensembles increase. Transparency is not a complete solution either as the results of displaying multiple transparent surfaces is dependent on the order of rendering. In addition, blending of too many surfaces prevents the user from being able to determine what surfaces were blended at any given point. When dealing with ensemble output, transparency can be effectively used to show areas of agreement but can not effectively identify sub-regions where there is partial agreement between different sets of ensemble members. All of these problems that are made more severe by the fact that the size of any given output can easily range in multiple gigabytes and rendering is not an instantaneous operation, particularly when done in software. These same problems also occur when trying to analyze multiple outputs describing different historical or statistical events.

*e-mail: dwj12@msstate.edu
†e-mail:tjk@acm.org

2 Approach

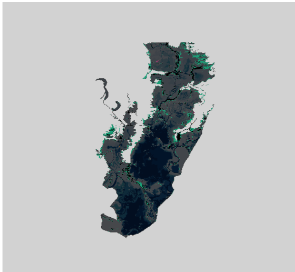
One of the major difficulties with processing geo-spatial data-sets is the potential size, and thus required processing time required for any input. This difficult is increased with each additional datase,t that must be processed, before results can be visualized or calculated. In order to reduce the necessary processing time to the point where an interactive system is achievable a data streaming architecture is used where data is processing is done on the gpu without the size of all input data being limited by GPU memory size.

The visualization system is designed along the overview + details on demand pattern. The overview display shows an accumulation mapping of the inputs into the display system. This means that the user is initially presented with an image showing how many inputs where defined (for some logical predicate) at each pixel location in the visualized domain. The overview display also shows a histogram that shows the counts of each level of overlap that is displayed. The details mode of the visualization system allows the use to see which combinations of surfaces exist for any given level of detail, and the spatial positions of each group of surfaces.

The visualization system works in overview by streaming individual surfaces to the GPU where they are aggregated into the accumulation map for user display. In details mode instead of accumulating a count of surfaces the GPU instead accumulates binary flag patterns. The resulting sums are then counted using a computational kernel and then the resulting pairs of count and value are returned to the CPU, where each value can used to determine the exact surfaces that where combined to achieve a given value. This allows the system to determine which of the possible surface combinations at a given level of overlap exist and then assign categorical colors to the the existing surface groups based on the number of recorded locations for each group. Categiorical color is only assigned to the first 10 surface groups, however surface groups beyond the first 10 do not tend to be of significant size.

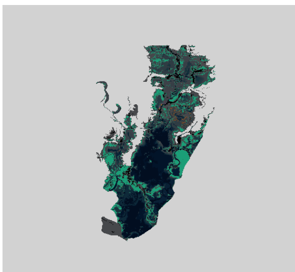
3 Results

The visualization system was tested with three different use cases. The first two use the system for analysis of multiple flood images and the last instead works with predictions from a weather model simulating the 1993 super storm.

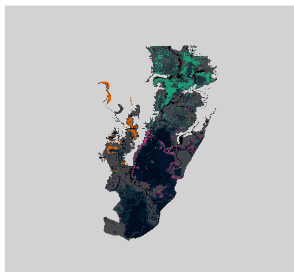


(a) Areas unique to the 25-year frequency flood

(b) areas common to only the 10 and 25-year frequency floods.



(c) overlap of three layers (5, 10 and 25-year frequency floods.



(d) the overlap of the 2, 5, 10, and 25-year frequency floods.

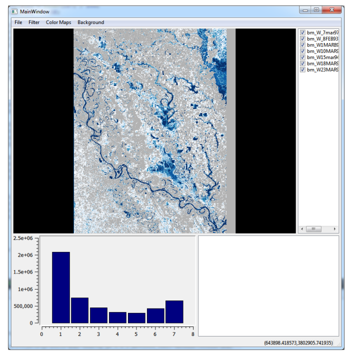
Figure 1: Flooding in the upper regions of the project area occurs only during frequency events, this indicates short term flooding only. The green surface shown in (a) is made from the 25 year frequency flood. In (b) the surface is the combination of the 25 and 10 year frequency floods. In (c) and (d) the green surface represents the combination of the 25, 10, and 5 year frequencies, and the 25, 10, 5 and 2 year frequencies respectively. Because all of the indicated green locations are covered only by frequency events it means none of them are ever flooded for a period of at least 14 days.

3.1 Yazoo Backwater

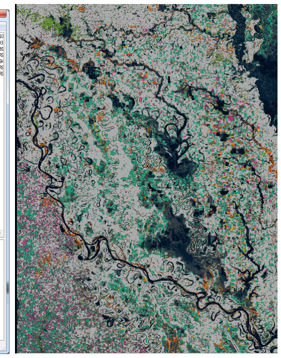
The Yazoo Backwater an approximately 1,550 square mile region between the eastern main line levee of the Mississippi River and the western Will Whittington Canal levee, bounded on the north US Highway 82. This area has been the focus of many hydrological studies performed by the USACE (Army Corps of Engineers). After removing the 50 year flood, (due to much large size) selecting the first few levels of overlap shows that all flooding in the northern area of region occurs only in frequency events (Figure 1); this is revealed by only surface interaction made up of frequency flood events covering the entire area. This means that none of the indicated areas are ever flooded for a period of two weeks in the events considered.

3.2 Bayou Meto

The Bayou Meto Wildlife Management Area is located off the Arkansas river southeast of Little Rock, AR and east of Pine Bluff, AR. Seven remotely sensed flood maps from this area were used to test the visualization system with satellite imagery, instead of the model outputs used previously. The initial visualization for this data set appears to be quite noisy (Figure 2). This can partially be explained by the fact farmers in the region purposefully flood some of their fields during waterfowl season; however, the particular fields artificially flooded changes year-to-year, resulting in more or less random areas of isolated flooding each scene. However, some insight can still be retrieved. The main channel of the arkansas river can easily be identified as well as various lakes and wetland regions by looking at areas of high overlap.



(a) Initial display for the Beyou Meto dataset



(b) Areas flooded in only one event

Figure 2: The initial display for the Beyou Meto dataset is shown in (a). The dominating class is surfaces that overlap with no other events. The surface break down for none overlapping surfaces is shown in (b).

3.3 1993 SupterStorm

The ensemble data we examine came from weather predictions of the 1993 “Superstorm” [2], a weather pattern that during its 3 day existence affected Central America, the Untied States, and Canada; it is one of the first examples of such a wide spread and sudden weather system being predicted by computational models. The data we examine came from a 30 member ensemble running the WRF model. There are two problems with the usage of this data set. First, the spatial resolution is poor—the actual data grids are only 129×157 even though the dataset itself measures over 61 Gigabytes in size. The second problem is that there is no information about the parameters settings in the different ensemble members. This means that although patterns of agreement and discord between ensemble members can be noted, we cannot discuss the source of these factors. In general, exploration of this data set reveals both areas of almost complete agreement or disagreement among the members of the ensemble. Disagreement, or uncertainty, in the result, was limited to the edges of the prediction areas as further explored in [2] but illustrated here.

4 Conclusion

We have presented a new GPU-accelerated visualization system that combines overviews, region selection, and statistical data to allow rapid comparison of multiple surface agreement and conflict. This system was tested with both ensemble data and data sets consisting of multiple flood scenes. Interacting with the visualization system allows features to quickly be located in both data types. Initial consultations with field experts in flood modeling indicate that the proposed system would greatly improve analysis of such datasets, and could save significant amounts of both time and money.

References

[1] N. Elmqvist, A. Vande Moere, H.-C. Jetter, D. Cernea, H. Reiterer, and T. J. Jankun-Kelly. Fluid interaction for information visualization. *Information Visualization*, 10(4):327–340, 2011.

[2] J. Sanyal, S. Zhang, J. Dyer, A. Mercer, P. Amburn, and R. Moorhead. Noodles: A tool for visualization of numerical weather model ensemble uncertainty. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1421–1430, Nov 2010.