**Proposal: Visual Exploration of Multi-faceted Halo Data Sets**

Xiaotong Liu

Meraj Khan

2014.3.28

**1. Introduction**

Cosmological simulations follow the formation of nonlinear structure in dark and luminous matter. In the Dark Sky Simulations project, each simulation is typically evolved over 14 billion years from very small fluctuations in an otherwise uniform distribution, using gravitational N-body integrators in an expanding background Universe. The particles in cosmological simulations cluster into gravitationally bound structures that pull in baryonic matter that form stars, galaxies, and clusters of galaxies.

One fundamental research topic in cosmological simulation is the identification and visualization of halos are formed by particles and defined by mass, position, and velocity and a set of additional properties of the halo, such as angular momentum, characteristic radius, spin parameter. These multi-faceted attributes complicate the identification of halos in such high dimensional data space. Visualizing and understanding multi-dimensional halo data set that is large in both size and dimensionality is a non-trivial task.

Our goal of this work is to design a visualization framework that assists users to analyze the multi-faceted properties of the halos for identification and visualization of halos in cosmological simulations. A multi-faceted view of halo properties can reveal the relationships between the facets and prompt selecting particles that are associated with halos of interest. We propose to combine multi-dimensional visualization techniques and particle visualization to support visual exploration of multi-dimensional halo data sets. We expect to provide linked views of the halos as well as their multi-faceted properties. Various interactions such as zooming and filtering will also be supported to allow users to explore the subspace of the data sets.

**2. Data set**

In this work, we use the IEEE VIS contest data set in 2015, which consists of three primary data types: the raw particle data, the Halo catalog data, and the merger tree data.

First, the raw particle data contains a position vector, velocity vector, and unique particle identifier. There are approximately 100 time steps, each stored in the SDF. This format is composed of a human readable ASCII header followed by raw binary data.

Second, the Halo Catalog data groups sets of gravitationally bound particles together into coherent structures. It has the information about a given halo’s position, shape, and size, are a number of statistics derived from the particle distribution, such as angular momentum, relative concentration of the particles, and many more. These catalogs are stored in both ASCII and binary formats.

Third, the Merger Tree data links the individual halo catalogs that each represents a snapshot in time, thereby creating database. Each merger tree forms a sparse graph that can then be analyzed to use quantities such as halo mass accretion and merger history to inform how galaxies form and evolve through cosmic time. Merger tree data are also distributed in both ASCII and BINARY formats.

**3. Tasks**

The main task to address in this work is to analyze the multi-faceted properties of the halos for identification and visualization of halos in cosmological simulations. A multi-faceted view of halo properties can reveal the relationships between the facets and prompt selecting particles that are associated with halos of interest. We will process the Halo Catalog data to visualize the multi-faceted properties of halos, and identify the major properties and ranges that define typical halo structures. Because there are a large number of properties in the halo data sets, selecting the informative and representative properties that can describe the halos is an important and challenging task. This *property view* will be linked with the *halo view* of halos, which provides users with the most direct and intuitive view of the halo structures. Users can manipulate the property view highlight halos of interest and examine the corresponding halo structures in the halo view.

**4. Software**

**Yt.py** is used to read and process the raw particle data in SDF format.

**D3.js** is used to create the PCP-based property view, and support interactions such as brushing and linking.

**Webgl.js** is used to render the halos in the halo view, and support interactions changing camera position and orientation of the view, and controlling the color and opacity of the halos.

We will create a web-based system that assists users to explore multi-faceted halo data sets.

**5. Visualization Design**

We propose to combine multi-dimensional visualization techniques and particle visualization to support visual exploration of multi-dimensional halo data sets. To this end, we employ parallel coordinates plot (PCP) to visualize the multi-faceted halo properties in the *property view*. PCP is a common way of visualizing high-dimensional data, representing a point in multiple-dimensional space as a polyline with vertices on parallel axes, where the position of the vertex on an axis corresponds to the value of that dimension. In our design, each axis corresponds to one property of the halo, and the point at which the polyline intersects represents the value of the point in that property. Because there are a large number of properties in the halo data sets, selecting the informative and representative properties that can describe the halos is important and challenging. We will resort to the correlations revealed from the PCP to filter out the relatively redundant dimensions of properties and focus on showing the informative ones. Besides, the particles that compose a halo are shown in the *halo view* using point-based visualization. These two views will be linked together to support the exploration process: users can brush the PCP in the property view to select halos of interest in a subspace, and examine the corresponding halo structures in the halo view.

**6. Expected Results and Evaluations**

We have downloaded the entire IEEE VIS contest data in 2015, and process the raw particle data using yt.py. We expect to provide linked views of the halos as well as their multi-faceted properties. Various interactions such as zooming and filtering will also be supported to allow users to explore the subspace of the data sets. To demonstrate the effectiveness and usefulness of our approach, we will perform a case study on the IEEE VIS contest data set in 2015 using our visualization system.