FAKULTÄT FÜR !NFORMATIK
Faculty of Informatics

Diplomarbeitspräsentation

Visualizing High-Dimensional Data with Hierarchically Aggregated Subsets

David Pfahler

TU Wien
Institute of Visual Computing and Human-Centered Technology
Research Division of Computer Graphics
Advisor: Ao.Univ.Prof. Univ.-Doz. Dipl.-Ing. Dr.techn. Eduard Gröller
Assistance: Dipl.-Ing. Dr.techn. Harald Piringer

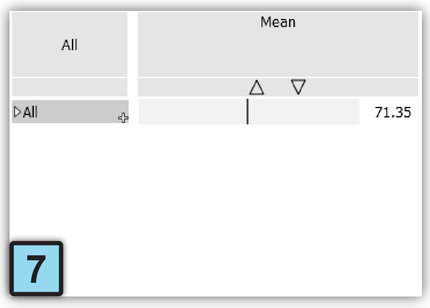
Master:
Visual Computing

Problem Statement / Motivation

The number of installed sensors to acquire data, for example electricity meters in smart grids, is increasing rapidly. The huge amount of collected data needs to be analyzed and monitored by transmission-system operators. This task is supported by visual analytics techniques, but traditional multi-dimensional data visualization techniques do not scale very well for high-dimensional data.

Main Contribution

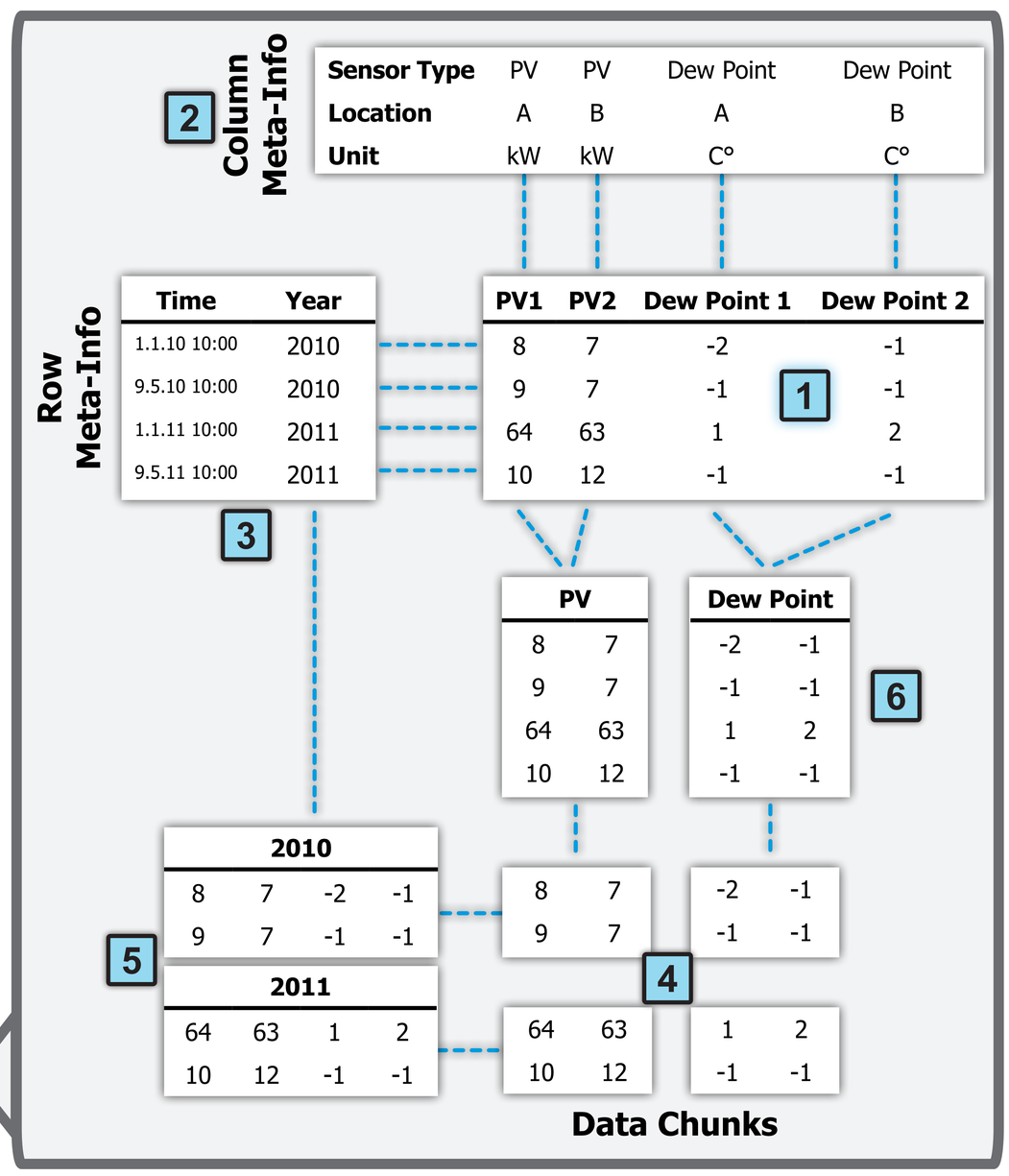
The main contribution of this thesis is a framework to efficiently examine and compare such high-dimensional data. The key idea is to divide the data by the semantics of the underlying dimensions into groups. Domain experts are familiar with the meta-information of the data and are able to structure these groups into a hierarchy. Various statistical properties are calculated from the subdivided data. These are then visualized by the proposed system using appropriate means. The hierarchy and the visualizations of the calculated statistical values are displayed in a tabular layout. The rows contain the subdivided data and the columns visualize their statistics.



1. Data table
2. Meta information on the columns
3. Meta information on the rows.

4. Subdivided data chunks
5. Splits of the rows
6. Splits of the columns

7. Statistical properties like the mean are calculated for the data table
8. Splitting the table into data chunks shows the mean values for every chunk.
9. Combining chunks by their meta information gives an overview
10. Subdividing the data further shows details for interesting aspects of the data.



Visualiziations of statistical properties of data chunks

Central tendencies

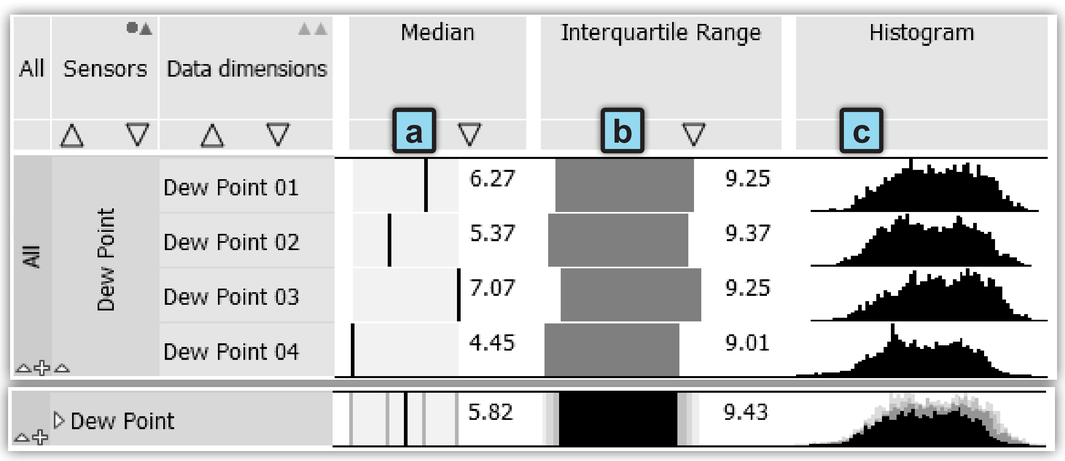
- a shown as a line on the horizontal position of its scale
- d shown as a line on its vertical position

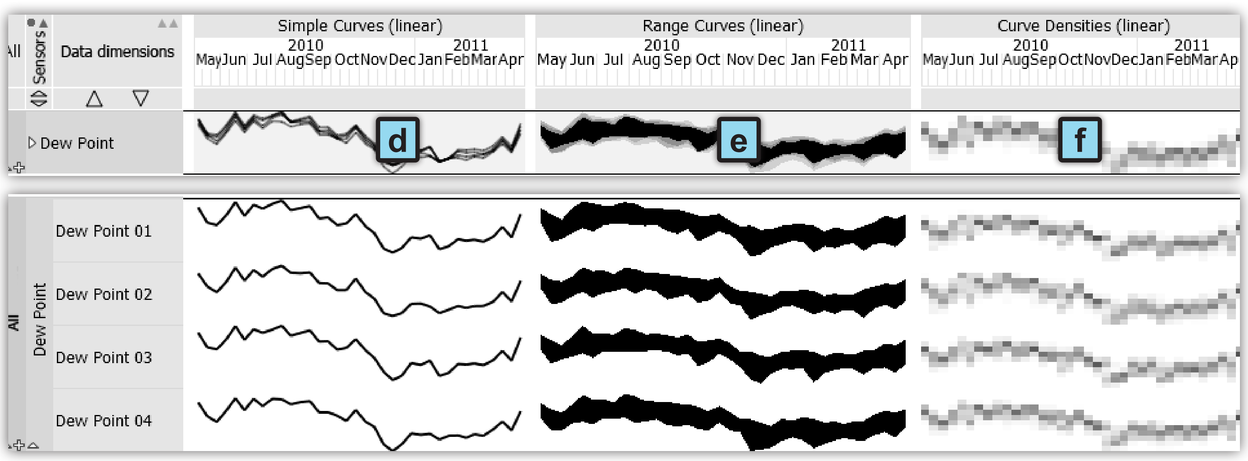
Dispersions

- b shown as an area around its corresponding central tendency
- e show an area graph

Frequency distributions

- c visualized by a histogram
- f shown as a heat map





Contact: pfahler@vrvis.at