GMaVis: A Domain-Specific Language for Large-Scale Geospatial Data Visualization Supporting Multi-core Parallelism

Cleverson Ledur
Advisor: Ph.D. Luiz Gustavo Fernandes
Co-Advisor: Ph.D. Isabel Manssour

Pontifical Catholic University of Rio Grande do Sul - PUCRS
Computer Science Graduate Program - PPGCC
Grupo de Modelagem de Aplicações Paralelas - GMAP

November 2015
Outline

1. Introduction
2. Related Work
3. GMaVis: A DSL for Geospatial Data Visualization
4. GMaVis’ Compiler
5. Evaluation and Discussion
6. Conclusion
1 Introduction
   • Contextualization
   • GMaVis Introduction

2 Related Work

3 GMaVis: A DSL for Geospatial Data Visualization

4 GMaVis’ Compiler

5 Evaluation and Discussion

6 Conclusion
Data visualization is the representation of data using graphic elements.

Provide a quick understanding of data.

Visualization creation pipeline has three steps: data pre-processing, data to visual mappings, and view transformations.

**Figure:** Visualization pipeline (Adapted from [1])
Users with low-level knowledge in programming that need to create a geo-spatial data visualization may have a hard time.

For example, to create a simple web visualization map, they will need to know at least JavaScript and HTML.

If they are dealing with a huge volume of data, it will be more difficult since most libraries and tools do not provide big data preprocessing.

It is possible to use parallel processing to speed up visualization creation.
**Parallel Programming**

- Developers must worry about:
  - Architecture System?
  - Parallel Interface?
  - Strategy: Shared Memory, Message Passing, Hybrid.
  - Decomposition of the problem: Data Parallelism, Task Parallelism, Stream Parallelism.
  - Steps: Study problem or code, Look for parallelism opportunities, Try to keep all cores busy.
  - Developer must care about Synchronization (Locks/Semaphores, Synchronous through communication, Barriers), Data Dependencies, Granularity, I/O, ...
1 Introduction
   ● Contextualization
   ● GMaVis Introduction

2 Related Work

3 GMaVis: A DSL for Geospatial Data Visualization

4 GMaVis’ Compiler

5 Evaluation and Discussion

6 Conclusion
We propose an external *domain-specific language* for the creation of large-scale data visualization focused on web data visualization map applications. The main goal is to provide a **high-level interface** that supports the visualization of the detail specification and the **manipulation of raw data automatically**, using a **parallel data pre-processor**.
**Introduction**

**Figure:** Research scenario framework (Extracted from [4]).
1 Introduction

2 Related Work
   - DSLs for Visualization

3 GMaVis: A DSL for Geospatial Data Visualization

4 GMaVis’ Compiler

5 Evaluation and Discussion

6 Conclusion
## Related Work

### DSLs for Visualization

<table>
<thead>
<tr>
<th>DSL</th>
<th>Domain</th>
<th>Focus</th>
<th>Parallelism</th>
<th>Interface</th>
</tr>
</thead>
<tbody>
<tr>
<td>ViSlang</td>
<td>Sci/Vol Vis.</td>
<td>Vol Rend.</td>
<td>CPU/GPU</td>
<td>High-level (H:C++)</td>
</tr>
<tr>
<td>Diderot</td>
<td>Img Analysis &amp; Med. Vis.</td>
<td>Img Rend. &amp; Analysis</td>
<td>CPU/GPU</td>
<td>High-level (H:C)</td>
</tr>
<tr>
<td>Shadie</td>
<td>Med Vis.</td>
<td>Vol Rend.</td>
<td>CPU/GPU</td>
<td>High-level (H:Phyton)</td>
</tr>
<tr>
<td>Superconductor</td>
<td>General Interactive Vis.</td>
<td>Rend.</td>
<td>CPU/GPU</td>
<td>High-level (Ext.)</td>
</tr>
</tbody>
</table>

- **Vivaldi**, **ViSlang**, **Diderot**, and **Shadie** are focusing in the generation of volumetric data visualizations. **Superconductor** allows the user to create maps because have more expressiveness, but it requires programming skills.

- **Google Maps API**, **Leaflet** and **OpenLayers** are visualization maps libraries with high-level abstractions. However, they do not avoid to learn a programming language, pre-processes and insert data.
Related Work

**Figure:** Data Visualization Creation Comparison
**Table:** Complexities abstraction in each visualization creation phase.

<table>
<thead>
<tr>
<th>DSL</th>
<th>Data Pre-processing</th>
<th>Data to Visual Mappings</th>
<th>View Transformations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vivaldi [3]</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>ViSlang [8]</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Diderot [2]</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Shadie [5]</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Superconductor [7]</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>GMaVis [6]</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

**Table:** Parallel processing in each visualization creation phase.

<table>
<thead>
<tr>
<th>DSL</th>
<th>Data Pre-processing</th>
<th>Data to Visual Mappings</th>
<th>View Transformations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vivaldi [3]</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>ViSlang [8]</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Diderot [2]</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Shadie [5]</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Superconductor [7]</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>GMaVis [6]</td>
<td>X</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
1 Introduction

2 Related Work

3 GMaVis: A DSL for Geospatial Data Visualization
   - GMaVis

4 GMaVis’ Compiler

5 Evaluation and Discussion

6 Conclusion
Proposed DSL

- Aim to **facilitate the creation** of visualizations.
- To be as close as possible **closer to the domain vocabulary** (supporting a suitable and friendly language syntax).
- **Users will not have to know programming** aspects like functions, variables, methods or any other web development issue.
- Users will have **automatic data processing** that empowers the data filtering, cleaning and, classification.
- **Optimized file loading in memory** that allows users to open files bigger than the RAM memory available in the system.
Architecture Overview

**Figure:** DSL Environment

**Figure:** Data Pre-processor Workflow
GMaVis

Interface - Elements

- High-level interface
- Few lines of code
- External DSL
- Manipulation and preprocessing of data

This DSL language rule set consists of blocks and declarations.

**Figure:** Example of Block and Declaration
Interface - Logical Operators

**Figure:** Logical Operators for Filtering and Classifying
**Interface Example - Traffic Accidents in Porto Alegre (Brazil)**

```
visualization: heatmap;

settings {
    latitude: field 42;
    longitude: field 41;
    page-title: "Acidentes Poa 2013";
    size: full;
    zoom-level: 11;
}

data {
    h-file: "acidentes-2013.csv";
    structure {
        delimiter: ';;'
        end-register: newline;
    }
}
```

**Figure:** Traffic accidents in Porto Alegre.
GMaVis

Interface Example - Flickr Pictures Classified by Brand of Used Camera

```plaintext
visualization: markedmap;
settings {
    latitude: field 12;
    longitude: field 11;
    marker-text: "<img src=" field 15 " width=200>";
    page-title: "Photos of 2014 by Camera";
    size: full;
}
data {
    file: "BIGDATA_YAHOO/yfcc100m_dataset-0";
    structure {
        delimiter: tab;
        end-register: newline;
        date-format: "YYYY-MM-DD";
    }
    filter: field 4 is greater than date "2014-02-01";
    classification {
        class ("Canon"): field 6 contains "Canon";
        class ("Sony"): field 6 contains "Sony";
        class ("Nikon"): field 6 contains "Nikon";
        class ("Panasonic"): field 6 contains "Panasonic";
        class ("Apple"): field 6 contains "Apple";
        class ("Kodak"): field 6 contains "Kodak";
    }
}
```

**Figure:** Flickr pictures by brand of camera used.
1 Introduction

2 Related Work

3 GMaVis: A DSL for Geospatial Data Visualization

4 GMaVis’ Compiler
   • Compiler Environment
   • Data Preprocessor Generator
   • Parallel Data Preprocessor

5 Evaluation and Discussion

6 Conclusion
Flex and Bison

Receives the source code and performs lexical and syntax analysis.

Generates tokens and joins them according to the specified grammar rules.

Each rule has different actions, such as: saving information, calling functions to process values, concatenation of strings and flagging.

**Figure:** GMaVis compiler environment.
1 Introduction

2 Related Work

3 GMaVis: A DSL for Geospatial Data Visualization

4 GMaVis’ Compiler
   - Compiler Environment
   - Data Preprocessor Generator
   - Parallel Data Preprocessor

5 Evaluation and Discussion

6 Conclusion
Data Preprocessor Generator

**Figure:** Example of logical expression transformation.

```
field 2 is equal to 1000
int_is_equal(counter, data, i, 2, 1000)
```

**Figure:** Data preprocessor generation.
### Data Preprocessor Generator

**Table:** Data preprocessor functions for each logical operator.

<table>
<thead>
<tr>
<th>Gmavis Logical Operator</th>
<th>Input Data Type</th>
<th>Generated Function in Data Preprocessor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is equal to</td>
<td>string</td>
<td>string.is_equal(counter, data, i, field, value)</td>
</tr>
<tr>
<td></td>
<td>date</td>
<td>date.is_equal(counter, data, i, field, date2)</td>
</tr>
<tr>
<td></td>
<td>integer</td>
<td>int.is_equal(counter, data, i, field, value)</td>
</tr>
<tr>
<td></td>
<td>float</td>
<td>float.is_equal(counter, data, i, field, value)</td>
</tr>
<tr>
<td>Is different than</td>
<td>string</td>
<td>string.is_different(counter, data, i, field, value)</td>
</tr>
<tr>
<td></td>
<td>date</td>
<td>date.is_different(counter, data, i, field, date2)</td>
</tr>
<tr>
<td></td>
<td>integer</td>
<td>int.is_different(counter, data, i, field, value)</td>
</tr>
<tr>
<td></td>
<td>float</td>
<td>float.is_different(counter, data, i, field, value)</td>
</tr>
<tr>
<td>Is greater than</td>
<td>date</td>
<td>date.is_greater(counter, data, i, field, date2)</td>
</tr>
<tr>
<td></td>
<td>integer</td>
<td>int.is_greater(counter, data, i, field, value)</td>
</tr>
<tr>
<td></td>
<td>float</td>
<td>float.is_greater(counter, data, i, field, value)</td>
</tr>
<tr>
<td>Is less than</td>
<td>date</td>
<td>date.is_less(counter, data, i, field, date2)</td>
</tr>
<tr>
<td></td>
<td>integer</td>
<td>int.is_less(counter, data, i, field, value)</td>
</tr>
<tr>
<td></td>
<td>float</td>
<td>float.is_less(counter, data, i, field, value)</td>
</tr>
<tr>
<td>Is between and contains</td>
<td>date</td>
<td>date.is_between(counter, data, i, field, date2, date3)</td>
</tr>
<tr>
<td></td>
<td>integer</td>
<td>int.is_between(counter, data, i, field, value1, value2)</td>
</tr>
<tr>
<td></td>
<td>float</td>
<td>float.is_between(counter, data, i, field, value1, value2)</td>
</tr>
<tr>
<td></td>
<td>string</td>
<td>string.contains(counter, data, i, field, value)</td>
</tr>
</tbody>
</table>
**Figure:** Data visualization generation.
1 Introduction

2 Related Work

3 GMaVis: A DSL for Geospatial Data Visualization

4 GMaVis’ Compiler
   - Compiler Environment
   - Data Preprocessor Generator
   - Parallel Data Preprocessor

5 Evaluation and Discussion

6 Conclusion
### Parallel Data Preprocessor

- Data preprocessor module has a **high computational cost** [6].
- We implemented a parallel version for **multi-core** architectures.
- The code that we implement to parallelize the data preprocessor using **SPar** is almost the same as is used in the sequential version.
- GMaVis compiler **generates SPar annotations**.
Figure: Example of sequential processing in data preprocessor.
Parallel Data Preprocessor

Figure: Parallel preprocessor using pipeline strategy.
```c
void process(...) {
    variables declaration
    calculate crunch size
    open file
    while(last_read < file_size)
        read a chunk from the file.
        verify its consistency
        start processing chunk
        returns a string with the output
        writes the output string in the output file
        deallocates memory for chunk read
    close file
}
```

**Figure:** Illustration of sequential process function code.
void process(...)
{
    variables declaration
calculate crunch size
open file
    [[spar::ToStream(), spar::Input(...)]]
while (last_read < file_size)
{
    read a chunk from the file.
verify data consistency
    [[spar::Stage(), spar::Input(...), spar::Output(...), spar::Replicate(...)]]
    start processing chunk
returns a string with the output
}
    [[spar::Stage(), spar::Input(...)]]
    writes the output string in the output file
deallocates memory for chunk read
}
close file

Figure: Illustration of process function code with SP\text{Ar} annotations.
Introduction

Related Work

GMaVis: A DSL for Geospatial Data Visualization

GMaVis’ Compiler

Evaluation and Discussion

Methodology
- Programming Effort Results
- Performance Results

Conclusion
## Research Questions

- **Q1** - Is it possible to create a DSL (GMAVIS) to reduce the programming effort for creating geospatial visualization?

- **H1** - GMaVis requires less programming effort than visualization libraries (Google Maps API, Leaflet and OpenLayers).

- **H2** - The inclusion of automatic data pre-processing in GMaVis reduces programming effort.

- **Q2** - Can the parallel code generated by this DSL speed up the data processing of raw geospatial data?

- **H3** - GMaVis can generate parallel code annotations using SPar for speeding up performance.

- **H4** - Code generation for SPar is simpler than using TBB (Thread Building Blocks) for speeding up performance.
Research Questions

- **Q1** - Is it possible to create a DSL (GMaVIS) to reduce the programming effort for creating geospatial visualization?
  - **H1** - GMaVis requires less programming effort than visualization libraries (Google Maps API, Leaflet and OpenLayers).
Research Questions

- **Q1** - Is it possible to create a DSL (GMAVIS) to reduce the programming effort for creating geospatial visualization?
  - **H1** - GMAVis requires less programming effort than visualization libraries (Google Maps API, Leaflet and OpenLayers).
  - **H2** - The inclusion of automatic data pre-processing in GMAVis reduces programming effort.
Research Questions

- **Q1 - Is it possible to create a DSL (GMAVIS) to reduce the programming effort for creating geospatial visualization?**
  - **H1** - GMaVis requires less programming effort than visualization libraries (Google Maps API, Leaflet and OpenLayers).
  - **H2** - The inclusion of automatic data pre-processing in GMaVis reduces programming effort.

- **Q2 - Can the parallel code generated by this DSL speed up the data processing of raw geospatial data?**
Research Questions

• Q1 - Is it possible to create a DSL (GMAVIS) to reduce the programming effort for creating geospatial visualization?
  • H1 - GMaVis requires less programming effort than visualization libraries (Google Maps API, Leaflet and OpenLayers).
  • H2 - The inclusion of automatic data pre-processing in GMaVis reduces programming effort.

• Q2 - Can the parallel code generated by this DSL speed up the data processing of raw geospatial data?
  • H3 - GMaVis can generate parallel code annotations using SPAR for speeding up performance.
Research Questions

- **Q1 - Is it possible to create a DSL (GMAVIS) to reduce the programming effort for creating geospatial visualization?**
  - **H1** - GMaVis requires less programming effort than visualization libraries (Google Maps API, Leaflet and OpenLayers).
  - **H2** - The inclusion of automatic data pre-processing in GMaVis reduces programming effort.

- **Q2 - Can the parallel code generated by this DSL speed up the data processing of raw geospatial data?**
  - **H3** - GMaVis can generate parallel code annotations using SPar for speeding up performance.
  - **H4** - Code generation for SPar is simpler than using TBB (Thread Building Blocks) for speeding up performance.
Methodology

Applications - Datasets

- **YFCC100M**: Provided by Yahoo Labs. It has about 54GB of data, divided into ten files. This is a public multimedia data set with 99.3 million images and 0.7 million videos, all from Flickr and under Creative Commons licensing.

- **Traffic Accidents**: Also we used a dataset from DataPoa with traffic accident data in Porto Alegre, Brazil that contains information about the type of accidents, vehicles, date and time, level and location. It has 39 fields, including latitude and longitude, and about 20.937 registers.

- **Airports**: A dataset obtained in OpenFlights\(^1\) that contains all the airports in the world was used, with information about latitude and longitude, city, country and airport code. It has 8107 registers with 12 fields with information about the airports.

Six visualization applications were created using these datasets.

- Two applications for each type of visualization provided in GMaVis (MarkedMap, Heatmap, and Clusteredmap).

- GMaVis, Google Maps API, Leaflet and OpenLayers.

- C/C++ data preprocessor were created for Google Maps API, Leaflet and OpenLayers.
Evaluation

- Q1 - Programming effort evaluation
  - COCOMO Model (Sloccount)
- Q2 - Performance evaluation
  - Execution Time
  - Throughput
Methodology

Programming Effort Evaluation

- This Validates the effectiveness of facilitating visualization creation with GMaVis.
- This evaluation compares the effort to create a visualization using GMaVis and traditional tools.
- Traditional tools: Google Maps API, Leaflet and OpenLayers.
- 6 Applications (2 for each visualization type).
- In order to estimate data preprocessing, we measured a code developed in C++ that loads, processes data and output in the format required by the application.
Methodology

SLOCCount

- COCOMO model
  - Usability model for measuring code and estimation metrics.
  - Development time and effort based on the physical source lines of code (SLOC/KSLOC).
  - **Entire development cycle** for generating a visualization, including the initial process of planning, coding, testing, documenting and deploying it for users.

- SLOCCount tool
  - It is a software measurement tool, which counts the physical source lines of code (SLOC), ignoring empty lines and comments.
  - It also estimates development time, cost and effort based on the original Basic COCOMO model.
Performance Evaluation

- It evaluates the parallel data preprocessor to verify if it achieves better performance compared to the sequential version.
- Since TBB was also considered to be used to parallelize data preprocessor module because it offers stream parallelism support, we compared its application to SPar.
- Verifies if a parallelized TBB version presents better results than the automatic generated SPar version of data preprocessor.
Methodology

Performance Evaluation - Hardware/Software

Multi-core computer Blade Dell PowerEdge M610 with the following specification:

- Two processors Intel Xeon Six-Core E5645 2.4GHz Hyper-Threading
- 24 GBytes of memory.

Software:

- Operational System: Ubuntu 14.04 LTS
- C/C++ Compiler: G++ 5.2
- SPAr Compiler: CINCLE
Programming Effort Results

1. Introduction
2. Related Work
3. GMaVis: A DSL for Geospatial Data Visualization
4. GMaVis’ Compiler
5. Evaluation and Discussion
   - Methodology
   - Programming Effort Results
   - Performance Results
6. Conclusion
Programming Effort Results

**Effort Evaluation - Programming Effort (Entire development cycle).**
Programming Effort Results

Code productivity results

Table: SLOC (Physical Lines of Code) for each application.

<table>
<thead>
<tr>
<th>Application</th>
<th>GmaVis</th>
<th>Google Maps</th>
<th>OpenLayers</th>
<th>Leaflet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ctr-Airp</td>
<td>15</td>
<td>34</td>
<td>46</td>
<td>46</td>
</tr>
<tr>
<td>Ctr-Comp</td>
<td>17</td>
<td>34</td>
<td>28</td>
<td>39</td>
</tr>
<tr>
<td>Hm-Airp</td>
<td>15</td>
<td>42</td>
<td>29</td>
<td>24</td>
</tr>
<tr>
<td>Hm-Accid</td>
<td>17</td>
<td>41</td>
<td>34</td>
<td>24</td>
</tr>
<tr>
<td>Mm-Dev</td>
<td>25</td>
<td>34</td>
<td>22</td>
<td>27</td>
</tr>
<tr>
<td>Mm-Accid</td>
<td>21</td>
<td>43</td>
<td>24</td>
<td>27</td>
</tr>
</tbody>
</table>
### Programming Effort Results

**Lines of code to parallelize data preprocessor using SPar and TBB**

<table>
<thead>
<tr>
<th>Application</th>
<th>Sequential</th>
<th>SPAR</th>
<th>TBB</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SLOC</td>
<td>Code Increase (%)</td>
<td>SLOC</td>
</tr>
<tr>
<td>CM_CP</td>
<td>210</td>
<td>2,857</td>
<td>274</td>
</tr>
<tr>
<td>CM_AIR</td>
<td>199</td>
<td>3,015</td>
<td>263</td>
</tr>
<tr>
<td>HM_AC</td>
<td>210</td>
<td>2,857</td>
<td>274</td>
</tr>
<tr>
<td>HM_AIR</td>
<td>199</td>
<td>3,015</td>
<td>263</td>
</tr>
<tr>
<td>MM_DEV</td>
<td>224</td>
<td>2,678</td>
<td>288</td>
</tr>
<tr>
<td>MM_ACID</td>
<td>216</td>
<td>2,777</td>
<td>280</td>
</tr>
</tbody>
</table>
1. Introduction

2. Related Work

3. GMaVis: A DSL for Geospatial Data Visualization

4. GMaVis’ Compiler

5. Evaluation and Discussion
   - Methodology
   - Programming Effort Results
   - Performance Results

6. Conclusion
### Performance Results

**Sequential GMaVis Execution Times**

<table>
<thead>
<tr>
<th>Size</th>
<th>Data Pre-processing</th>
<th>Data to Visual Mappings - Google Maps API</th>
</tr>
</thead>
<tbody>
<tr>
<td>10GB</td>
<td>110.4948 (Std. 0.9763)</td>
<td>2.910 (Std. 1.6084)</td>
</tr>
<tr>
<td>50GB</td>
<td>544.0506 (Std. 9.4225)</td>
<td>3.2738 (Std. 2.0663)</td>
</tr>
<tr>
<td>100GB</td>
<td>1098.9284 (Std. 19.0383)</td>
<td>3.8536 (Std. 2.7584)</td>
</tr>
</tbody>
</table>

**Table:** Completion times (seconds) [6].
Performance Results

TBB and SPAR SLOC

Performance Results (Airports in world - Clusteredmap)

Performance Results (Flickr Photos with "Computer" Word as Tag - Clusteredmap)
Performance Results

TBB and SPAR SLOC
Performance Results

**TBB and SPAR SLOC**

### Performance Results (Traffic Accidents in Porto Alegre - Markedmap)

- **Big Dataset (Execution Time)**
  - TBB: [Graph]
  - SPAR: [Graph]

- **Medium Dataset (Execution Time)**
  - TBB: [Graph]
  - SPAR: [Graph]

- **Small Dataset (Execution Time)**
  - TBB: [Graph]
  - SPAR: [Graph]

- **Big Dataset (Throughput)**
  - TBB: [Graph]
  - SPAR: [Graph]

- **Medium Dataset (Throughput)**
  - TBB: [Graph]
  - SPAR: [Graph]

- **Small Dataset (Throughput)**
  - TBB: [Graph]
  - SPAR: [Graph]

### Performance Results (Flickr Photos by Device - Markedmap)

- **Big Dataset (Execution Time)**
  - TBB: [Graph]
  - SPAR: [Graph]

- **Medium Dataset (Execution Time)**
  - TBB: [Graph]
  - SPAR: [Graph]

- **Small Dataset (Execution Time)**
  - TBB: [Graph]
  - SPAR: [Graph]

- **Big Dataset (Throughput)**
  - TBB: [Graph]
  - SPAR: [Graph]

- **Medium Dataset (Throughput)**
  - TBB: [Graph]
  - SPAR: [Graph]

- **Small Dataset (Throughput)**
  - TBB: [Graph]
  - SPAR: [Graph]
1 Introduction

2 Related Work

3 GMaVis: A DSL for Geospatial Data Visualization

4 GMaVis’ Compiler

5 Evaluation and Discussion

6 Conclusion
Final Remarks

- We provided a new domain-specific language for the geospatial data with simpler and friendly interface.
- It offers support for big raw data sets.
- We used a real-world data set to evaluate our DSL, comparing it in different visualization types.
Final Remarks - Research Questions

Q1 - *Is it possible to create a DSL (GMAVIS) to reduce the programming effort for creating geospatial visualization?*

- Results demonstrate that even when providing the abstraction of complexities in the *data pre-processing* phase and totally parallel programming abstraction, GMaVis could reduce the effort and cost required to implement the three supported data visualizations (clusteredmap, heatmap, and markedmap) (*Validades H1*).

- It is possible to confirm that GMaVis can reduce not only the programming effort but also the cost of development for creating geospatial data visualizations (*i.e.*, markedmaps, clusteredmaps and heatmaps)(*Validades H2*).
Conclusion

Final Remarks - Research Questions

Q2 - *Can the parallel code generated by this DSL speed up the data processing of raw geospatial data?*

- The GMaVis compiler implementation demonstrates that it is possible to generate SPar annotations automatically to speed up the data processing (*Validades H3*).
- Results demonstrate that both versions (SPar and TBB) increased performance by decreasing the execution time when compared to the sequential version, but SPar required less code generation (*Validades H4*).
Shared memory architectures **have a bottleneck when reading data from disk**. First version of GMaVis was implemented targeting these types of architectures.
Conclusion

Limitations

- Shared memory architectures **have a bottleneck when reading data from disk**. First version of GMaVis was implemented targeting these types of architectures.

- **Low expressiveness**. This is because to offer a higher level of abstraction, it is required to keep some features hidden from users that allow changing some visualization details, such as colors, icons, and specific sizes.
Conclusion

Limitations

- Shared memory architectures have a bottleneck when reading data from disk. First version of GMaVis was implemented targeting these types of architectures.

- Low expressiveness. This is because to offer a higher level of abstraction, it is required to keep some features hidden from users that allow changing some visualization details, such as colors, icons, and specific sizes.

- This DSL has focused in geospatial data visualizations. However, data analysis may require the use of different visualization types/techniques to achieve its objective.
Limitations

- Shared memory architectures **have a bottleneck when reading data from disk**. First version of GMaVis was implemented targeting these types of architectures.

- **Low expressiveness**. This is because to offer a higher level of abstraction, it is required to keep some features hidden from users that allow changing some visualization details, such as colors, icons, and specific sizes.

- This DSL has **focused in geospatial data visualizations**. However, data analysis may require the use of different visualization types/techniques to achieve its objective.

- The data preprocessor in GMaVis only supports **non-hierarchical** data files.
Future Work

- Improve this DSL to perform in distributed memory architectures.
Future Work

- Improve this DSL to perform in **distributed memory architectures**.
- The insertion and support of **other visualization types** is planned for the future, using geospatial data.
Future Work

- Improve this DSL to perform in **distributed memory architectures**.
- The insertion and support of **other visualization types** is planned for the future, using geospatial data.
- The creation of a pre-parser for **hierarchical files** are planned for future work. Also, it is possible to use or convert a hierarchical into a non-hierarchical data file.
Future Work

- Improve this DSL to perform in distributed memory architectures.
- The insertion and support of other visualization types is planned for the future, using geospatial data.
- The creation of a pre-parser for hierarchical files are planned for future work. Also, it is possible to use or convert a hierarchical into a non-hierarchical data file.
- Data declarations can be extended in this interface to offer smart classifications and data selection through the use of machine learning and data mining algorithms.
Related Papers


**(SUBMISSION) Ledur, C.; Griebler, D.; Fernandes, L. G.; Manssour, I.** ”GMaVis: A Domain-Specific Language for Geospatial Data Visualizations”. In: IEEE Information Visualization (InfoVis), IEEE VIS, 2016
Published Papers

Contributions


References I

*Readings in Information Visualization: Using Vision to Think.*

C. Chiw, G. Kindlmann, J. Reppy, L. Samuels, and N. Seltzer.
Diderot: A Parallel DSL for Image Analysis and Visualization.
In *Proceedings of the ACM SIGPLAN Conference on Programming Language Design and Implementation*,
volume 47 of *PLDI’12*, pages 111–120, New York, USA, June 2012. ACM.

Vivaldi: A Domain-Specific Language for Volume Processing and Visualization on Distributed Heterogeneous Systems.

D. Griebler.
*A New Compiler-Based Framework Perspective for High-Level Stream Parallelism.*


Towards a Domain-Specific Language for Geospatial Data Visualization Maps with Big Data Sets.
In *ACS/IEEE International Conference on Computer Systems and Applications*, AICCSA’15, page 8,
Marrakech, Marrocos, November 2015. IEEE.
References II


P. Rautek, S. Bruckner, M. Groller, and M. Hadwiger.
ViSlang: A System for Interpreted Domain-Specific Languages for Scientific Visualization.
Conclusion

Questions

Questions & Answers
Thank you!
More questions:
cleverson.ledur@acad.pucrs.br
cleversonledur@gmail.com
http://gmap.pucrs.br/cleversonledur/