

# Generation of Semantic Layouts for Interactive Multidimensional Data Visualization

Erick Gomez-Nieto<sup>1</sup>, Luis Gustavo Nonato<sup>1</sup>

<sup>1</sup>ICMC – Universidade de Sao Paulo  
Sao Carlos, SP, Brazil.

**Abstract.** *Visualization methods make use of interactive graphical representations embedded on a display area in order to enable data exploration and analysis. These typically rely on geometric primitives for representing data or building more sophisticated representations to assist the visual analysis process. One of the most challenging tasks in this context is to determinate an optimal layout of these primitives which turns out to be effective and informative. However, most techniques are able to tackle just a few fundamental requirements simultaneously, impairing their use and flexibility. In this dissertation, we propose a set of approaches for building layouts from geometric primitives that concurrently addresses a wider range of requirements. A comprehensive set of quantitative comparisons against existing methods for layout generation and applications on multimedia data visualization prove the effectiveness of our approaches.*

Arranging geometric primitives to generate meaningful layouts is a major task in visualization, which inherently appears in important applications such as word cloud construction and visual boards. The difficulty in building layouts made up of dozens of geometric objects rests in the set of requirements to be handled simultaneously, e.g., readability, overlaps, object size, semantic proximity and area usage. Moreover, the number of data instances represented as geometric entities is typically much larger than the visualization area, demanding the use of clustering, hierarchies, and navigation resources to assist the visualization.

Although significant advances have been made towards building meaningful layouts from geometric primitives, existing techniques are formulated to deal with a limited number of requirements simultaneously, restricting their use to specific applications. For instance, techniques such as visual boards and small multiples provide well structured layouts which are easily readable, but they pay the price of scalability. Hierarchical methods such as Treemaps mitigate the issue of scalability while making an efficient use of display area. However, readability and semantic organization of data are aspects not so easily handled by those methods. Overlap-free semantic preserving techniques generate somewhat structured layouts and keep instances with similar content close to each other. Nevertheless, they are not designed to make an efficient use of display area and also suffer from scalability.

Handling many requirements is not straightforward because distinct requirements can compete with each other during layout construction. For instance, to facilitate readability, layouts should be built with as large as possible geometric entities. However, large objects easily fill up the display area, thus limiting the number of instances that can be visualized. Therefore, finding an optimal balance among multiple concurrent requirements is a challenging task, which has not been completely tackled by existing methods.

In this dissertation we addressed this challenging problem by proposing new techniques for building layouts. We denominate to *Semantic Layout Arrangement* as the task of allocating efficiently a set of geometric instances, which summarizes a multidimensional dataset, into a fixed-size display area, subject to preserve, as much as possible and at all times, the semantic relationships among instances. The arrangement should simultaneously play with several requirements, such as area usage optimization, overlap removal, object scaling, orthogonal alignment and dynamic updating.

This document presents a compilation of different techniques for interactive and semantic layouts generation for data visualization. Each proposed method brings new contributions for the field with the purpose of addressing and solving specific problems involved in generating geometric semantic layouts for interactive data visualization. Essentially, three proposed methods (*ProjSnippet* [Gomez-Nieto et al. 2014b], *MIOLA* [Gomez-Nieto et al. 2013] and *Dealing with Multiple Requirements* [Gomez-Nieto et al. 2016]) rely on novel optimization formulations for simultaneously dealing with requirements that were identified as the most relevant during our study. The last method (*Semantically Aware Dynamic Layouts* [Gomez-Nieto et al. 2014a]) focuses on an application that demands semantic preserving layout updates during analyst’s interactive data exploration.

To our knowledge, the proposed approaches figure among the most straightforward, efficient and intuitive alternatives to explore large amounts of multidimensional data while enabling a fluent interaction with data analysts.

## 1. ProjSnippet

The proposed technique comprises three steps: pre-processing of search results, multidimensional projection, and optimization. In the first step each entry returned from a textual query is processed and its term frequency vector extracted

Each term frequency vector may be handled as a point in a multidimensional space that can be mapped to the visual space with a multidimensional projection technique. Albeit our current implementation adopts the *Least Squares Projection (LSP)* – due to its good accuracy in terms of distance preservation and low computational cost – any projection technique with similar properties might be employed. The projection preserves much of the neighborhood structure of the original data, ensuring that similar instances are placed close to each other in the visual space.

The following step is to embed the content of each snippet within a rectangle whose bottom left corner is placed in the snippet’s (or its multidimensional data point) projected position. A rectangle’s height and width are settled to reflect the rank of its corresponding snippet in the retrieved document list, so that better ranked snippets are assigned larger rectangles.

A major drawback at this stage of the pipeline is that rectangles enclosing the snippets overlap considerably, impairing identification of individual entries and the perception of the document neighborhood structure. The final step optimizes the placement of the snippets in order to avoid overlapping while preserving data neighborhoods as computed by the projection.

The example illustrates a visualization displaying the results of a query on the

Jaguars, Jaguar Pictures, Jaguar Facts - National Geographic  
 www.nationalgeographic.com/mammals/jaguar/...  
 Learn all you wanted to know about jaguars with pictures, videos, facts, and...  
 Photo: A young female jaguar stepped in its tracks ... Big Cats Features ...

Jaguar International - XF Models and Features  
 www.jaguar.com/usa/xf...  
 Jaguar XF: A sporting saloon with a choice of four trim levels and six engines including 3.0 litre V6 Super S, first motorised about the XF, XF Portfolio and six...

Jaguar USA - FEATURES DISCLAIMER  
 www.jaguar.com/usa/features...  
 These Features are not a substitute for driving safely with due care and attention and will not function under all circumstances, speeds, weather and road ...

Jaguar USA - EXTERNAL FEATURES  
 www.jaguar.com/usa/features...  
 Jaguar XF: A sporting saloon with a choice of four trim levels and six engines including 3.0 litre V6 Super S, first motorised about the XF, XF Portfolio and six...  
 Feature a dramatic full-width glass roof, swinging back from the top of the ...

Jaguar International - XK Models and Features  
 www.jaguar.com/usa/xk...  
 Available in coupe or convertible bodystyles, the XK is available with a choice of two performance engines. First and more about XK, XK Portfolio and XKC.

Jaguar Specifications | Engine Performance & Jaguar Features ...  
 www.jaguar.com/usa/specifications...  
 The Motor Trend Road & Track's guide has detailed used Jaguar specifications, performance, standard equipment and comprehensive Jaguar data.

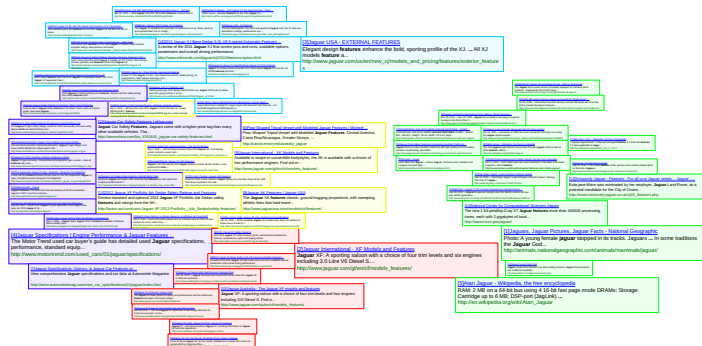
Asian Jaguar - Wikipedia, the free encyclopedia  
 en.wikipedia.org/wiki/Asian\_Jaguar  
 Jump to other Jaguar features: RAM: 2 MB on a 64-bit bus using a 16-bit fast page mode DRAM: Storage Cartridge - up to 4 MB; DSP port (Jaguar) ...

Used Jaguar Specifications, Options, & Jaguar Features at ...  
 www.autocheck.com/used...  
 View comprehensive used Jaguar specifications and car data at Automobile Magazine.

Pearl-Shaped Tripod Vase(s) with Model(s) Jaguar Features | Michael ...  
 www.etsy.com/listing...  
 Pearl-Shaped Tripod Vase(s) with Model(s) Jaguar Features. Central America, Costa Rica/Ricragua, Greater Mexico, Family Photographs, Family Vase(s), Pearl ...

Unusually Jaguar - Features: For all your Jaguar needs... Jaguar ...  
 www.unusuallyjaguar.co.uk/05\_Features.php  
 Can you believe it? Jaguar Land Rover, has a potential candidate for the City of Coventry Freeman's Guild Apprentice of the Year ...

(a) First page of Google



(b) ProjSnippet displaying 64 snippets

Figure 1. Google and ProjSnippet views of the results of a query with terms “jaguar features”. More here: <https://goo.gl/c4apaC>

terms “jaguar features” submitted to Google’s search engine. The view in Figure 1(a) shows the 10 best ranked snippets shown in the first page. Figure 1(b) displays a ProjSnippet view with the 64 best ranked snippets. Inspection discloses that the snippets on the left (cyan, red, blue, yellow) all refer to different models of Jaguar cars, whereas the green ones on the right refer to a surprising variety of topics, that include multiple references to the wild animal (3 snippets) and also to supercomputer models named Jaguar (2 instances). Still, overall the final layout depicts a representative overview of the search hits, as far grouping/separating similar/dissimilar results is concerned. Notice that it is pretty difficult to handle such a variety topics and subtopics in Google’s list-based view, which indeed brings only results on cars, animals and the game in the first page.

## 2. Mixed Integer Optimization for Layout Arrangement (MIOLA)

In this section we describe a technique to tackle the problem of arranging rectangular boxes in the visual space so as to place objects representing similar content close to each other while avoiding overlaps. We formulate the problem as a *Mixed Integer Quadratic Programming Problem* (MIQP), which enables well structured layouts. In contrast to other optimal methods that take into account the similarity between instances, our approach does not rely on intersection tests, making the algorithm simpler to implement. Moreover, our technique is quite flexible, being able to generate different layouts by just handling optimization constraints.

Let  $B = \{B_1, B_2, \dots, B_n\}$  be a set of  $n$  rectangular boxes arranged in the visual space such that the neighborhood structure of the boxes reflects a property of interest. For instance, if a data set is mapped to the visual space using a multidimensional projection technique and a box is centered on each projected data, the resulting arrangement makes neighbor boxes correspond to similar data. Boxes in this arrangement, however, should overlap considerably, impairing the visualization of individual boxes. In order to make each box visible, one has to displace the boxes in the two-dimensional space so as to remove overlaps, but preserving the initial neighborhood structures to keep similar objects

close to each other. As described in our paper, we formulate the problem above as a mixed integer quadratic programming optimization.

### 3. Dealing with multiple requirements for geometric layouts arrangement

In this section we describe a methodology for building layouts from geometric primitives which is able to deal with a wide range of requirements simultaneously. Relying on multi-dimensional projection, density-based adaptive grids, and mixed integer optimization, our approach is semantically aware, makes an efficient use of display area, and generates well structured grid-like layouts. Moreover, the formulation intrinsically impose a hierarchy on the data, enabling alternatives for the scalability issue.

The proposed optimization scheme arranges geometric entities (boxes) with varying sizes so as to avoid overlaps while preserving the neighborhood structure of the underlying data (semantics). The area of each geometric primitive is also included in the optimization process to ensure that the display area will be efficiently occupied. In fact, supported by the adaptive grid, our formulation is able to scale elements with different sizes using only one variable, thus rendering the optimization procedure as simple as possible.

Figure 2 depicts qualitative results comparing our approaches with overlap removal in 5 datasets. Notice that the proposed method (Our) gives rise to well structured layouts where neighborhoods (indicated by color map) are nicely preserved. Moreover, our approach makes a better use of display area, thus improving readability and content analysis.

### 4. Semantically Aware Dynamic Layouts

In this section we describe a novel semantic aware layout construction technique that allows users to freely tailor 2D arrangements according to their interest. The proposed formulation relies on interactive mechanisms enabled by multidimensional projection methods to enforce semantic relation in the layout. Moreover, the proposed approach is based on a simple energy function that can efficiently be minimized using well-known optimization libraries, thus avoiding intricate computational implementations.

The proposed mechanism to dynamically update layouts according to user intervention builds upon the methodology of *ProjSnippet*, which has been conceived to optimize layouts restricted to rectangular domains and with no interactive resource. Our approach, in contrast, enables interactive resources that allow users freely modify the layout according to their interest while still being able to build arrangements in arbitrary visual domains. Moreover, our formulation combines the flexibility provided by control points used in multidimensional projection with an energy function tuned to enable interactive layout update as well as to enforce semantic relation among neighbor entities.

Figure 3 shows layouts generated from fifty images with dimensions  $256 \times 256$  obtained from the Calltech 101 data set. Figure 3(a) shows the initial layout obtained by projecting and optimizing the layout using *ProjSnippet* (a square box has been assigned to each projected image). Figure 3(b) show the layout during user interaction phase. The overlap removal opens a “path” throughout the region the user has dragged images. The opened path helps the user to keep a mental model of how much and where the

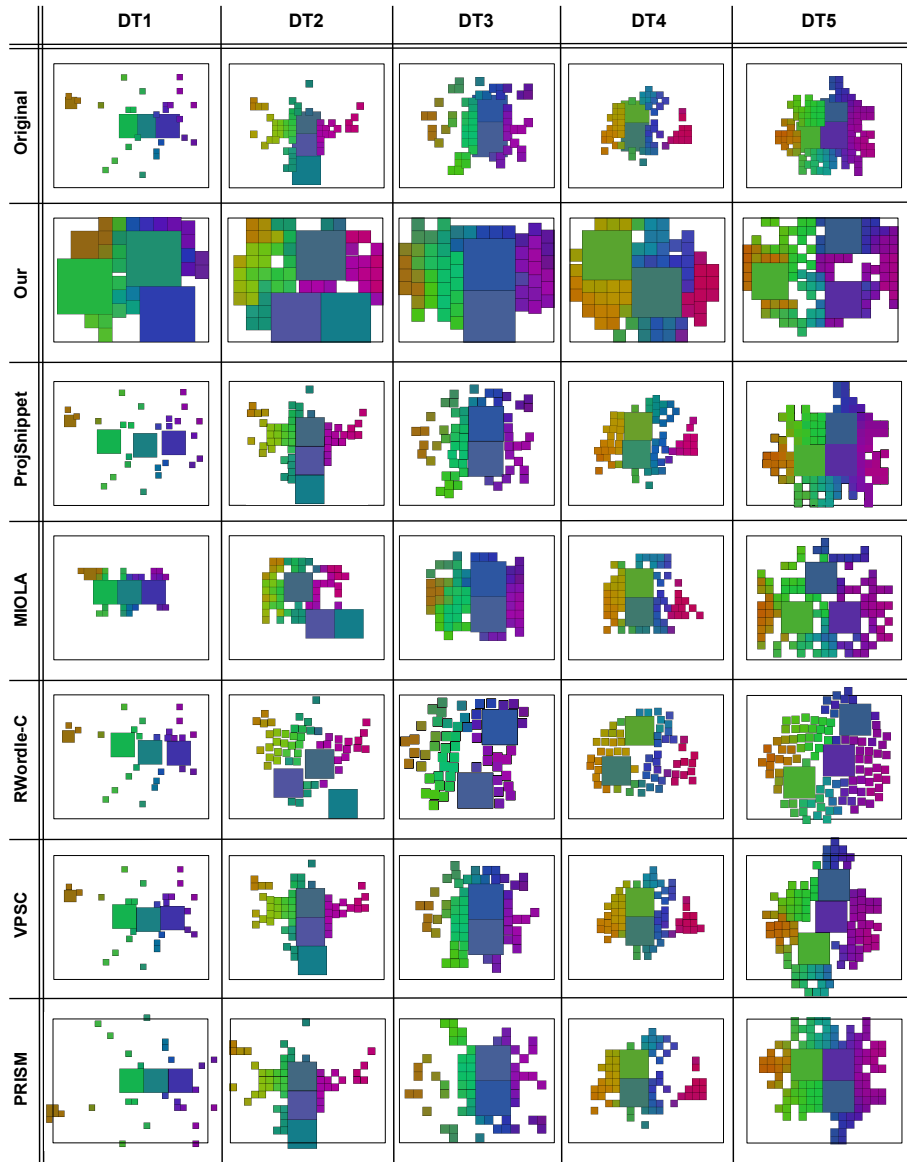


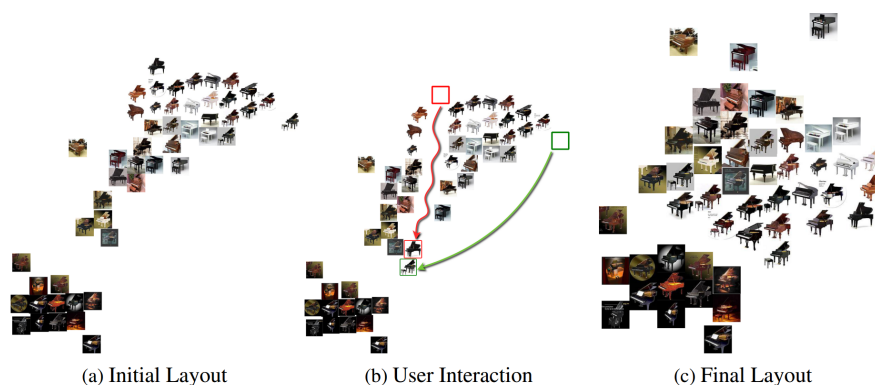
Figure 2. Layouts produced by our approaches and RWordle-C, VPSC and PRISM.

layout is changing. Figure 3(c) depicts the final layout after optimization. Notice that the interaction changed the position of two white background piano images and, as a consequence, all pianos with white background has also been properly displaced, showing the capability of our approach in preserving semantic relations.

## 5. Publications, Awards and Honors

### Publications

<i>Qualis A1</i>	<i>Qualis A2</i>	<i>Qualis B1</i>
[Gomez-Nieto et al. 2014b] [Gomez-Nieto et al. 2016]	[Zanabria et al. 2016]	[Gomez-Nieto et al. 2013] [Gomez-Nieto et al. 2014a] [Gomez-Nieto et al. 2015]



**Figure 3. Layouts produced from piano's images.** More here: <http://goo.gl/GEvFp2>

### Awards and Honors

- Best Computer Graphics/Visualization Ph.D. Thesis Award at SIBGRAPI 2017.
- Invited as PhD researcher intern at IBM Research, SP Lab (Mar/16-Feb/17).
- Invited IEEE TVCG paper presented at IEEE VIS 2016<sup>1</sup>.
- Accepted proposal to Doctoral Colloquium at IEEE VIS 2015.
- Invited IEEE TVCG paper presented at IEEE VIS 2014.
- Best paper award in Graphics and Visualization at SIBGRAPI 2013.

### References

- [Gomez-Nieto et al. 2015] Gomez-Nieto, E., Casaca, W., Hartmann, I., and Nonato, L. G. (2015). Understanding large legal datasets through visual analytics. In *Proceedings of 6th Workshop on Visual Analytics, Information Visualization and Scientific Visualization (WVIS) in SIBGRAPI'15*.
- [Gomez-Nieto et al. 2016] Gomez-Nieto, E., Casaca, W., Motta, D., Hartmann, I., Taubin, G., and Nonato, L. G. (2016). Dealing with multiple requirements in geometric arrangements. *IEEE Transactions on Visualization and Computer Graphics*, 22(3):1223–1235.
- [Gomez-Nieto et al. 2013] Gomez-Nieto, E., Casaca, W., Nonato, L. G., and Taubin, G. (2013). Mixed integer optimization for layout arrangement. In *Proceedings of the 26th Conference on Graphics, Patterns and Images (SIBGRAPI'13)*, pages 115–122.
- [Gomez-Nieto et al. 2014a] Gomez-Nieto, E., Motta, D., and Nonato, L. (2014a). Semantically aware dynamic layouts. In *Proceeding of the 27th Conference on Graphics, Patterns and Images (SIBGRAPI'14)*, pages 220–226.
- [Gomez-Nieto et al. 2014b] Gomez-Nieto, E., Roman, F. S., Pagliosa, P., Casaca, W., Helou, E. S., de Oliveira, M. C. F., and Nonato, L. G. (2014b). Similarity preserving snippet-based visualization of web search results. *IEEE Transactions on Visualization and Computer Graphics*, 20(3):457–470.
- [Zanabria et al. 2016] Zanabria, G. G., Nonato, L. G., and Gomez-Nieto, E. (2016). iStar (i\*): An interactive star coordinates approach for high-dimensional data exploration. *Computers & Graphics*, 60:107 – 118.

<sup>1</sup>IEEE VIS is the worldwide largest and most important conference on data visualization