Revisiting Landscape Views in Information Visualization[‡]

Pedro Henrique Siqueira* Guilherme P. Telles* Rosane Minghim[†]

* Instituto de Computação
Universidade Estadual de Campinas
opedrosiqueira@gmail.com, gpt@ic.unicamp.br

[†] Instituto de Ciências Matemáticas e de Computação
Universidade de São Paulo
rminghim@icmc.usp.br

Abstract

While the field of visualization has progressed immensely towards supporting data analysis tasks in the past decade, the discussion on the use of the third dimension in visual spaces for reflecting data has encountered both objections and support along the way. However, the advance of mapping techniques and interaction approaches have brought new light into the possibilities to use 3D views for cases where the number of dimensions actually makes a difference for the user. In that context, landscape views have too long being overlooked as a tool for exploration. While surfaces are mathematically bi-dimensional, the use of the third coordinate plane to express elements poses additional charges to the visual system, such as occlusion and challenges interacting with the layout. In this work we revisit landscape views in the light of recent advances in visualizations of high-dimensional data, which have provided better ways to create bi-dimensional mappings (such as evolved multidimensional scaling algorithms) and to summarize sets of data (such as word clouds and image mosaics). These advances provide additional clues to the user so that the landscape view is capable of both offering an overall view of the data set and suggesting entry points for further exploration of subspaces of interest. This article presents a number of strategies to accomplish a better surface data visualization and illustrates applications of the approach that reflect its potential use in real life cases. Our contribution includes methods to cluster and segment a projection, and recursive divisions of the visual space enriched by textures, which have various uses in multidimensional visualization and analysis tasks.

1 Introduction

Information visualization concerns building visual layouts of abstract data objects to unveil relationships and patterns within a dataset. In contrast with 3D object visualization (such as 3D for medical imagery and support for engineering), where 3D is a natural way to show target objects, abstract data have to be translated to a conceptual visual model before being presented to analysts.

For abstract data, many types of visual layouts have already been proposed that make use of both 2D and 3D views. Many of these tools, such as multidimensional projections, are capable of mapping data expressed as samples in a multidimensional space into visual spaces (2D or 3D) via transformations aimed at preserving some

[‡]PHS acknowledges the financial support of FAPESP (fellowship 2012/24634-6). GPT and RM acknowledge the financial support of FAPESP (grant 2011/22749-8) and CNPq.

property (such as distances or neighborhoods). We focus here on visually enriching point placements, that is, in defining a surface view as output for techniques that produce sets of points placed on a 2D plane. We are concerned with adding information to that layout in order to improve its expressiveness and to allow additional analysis to be done by means of a single view.

On top of a point based presentation, visual clues can be added to increase analysis power, such as color, texture, or iconographic displays for the points. A surface model can also be built by using the third dimension to reflect a particular variable or concept. That surface can reflect, through its shape, a global view of the dataset regarding the relationship between the distance distribution of the points and the particular concept mapped to the third coordinate. That visualization model is known as a landscape view or information landscape. Landscape views have been proposed many years ago, and some have been implemented in software tools that are currently in use, such as In-spire (`http://in-spire.pnnl.gov/about.stm`), a text visualization software. One of the problems with most currently available landscape views for abstract data is that the basis for landscape construction is subject to the uncertainty of the point placement strategy.

In point placements, distances between dots in 2D or 3D visual space are meant to reflect distance between data objects in original multidimensional space. Another spatial representation may be obtained by coloring the space among dots, producing a kind of heatmap that is a flat landscape. Variation in height associated with points or regions produce a surface embedded in 3D, with an additional degree of information given by the shapes formed in surfaces regions. Including visual artifacts such as color and icon shapes can assist exploration in both dot and landscape displays.

The effectiveness of any 3D views for data exploration has been widely debated. Advantages and disadvantages have been observed as will be recalled later. While the debate is still open, it is also illustrated by many recurring strategies that a good visual design and a simple mapping to the third dimension can help adding expression where a 2D view that is already cluttered or limited to convey, at once, general and particular views of the dataset. This last case, that is, adding expressive power for better exploration, is the most central point in this work.

The mappings shown in this work are actually on a surface, which for most purposes is a 2D element, however embedded in 3D. Visualizations of the resulting object may suffer from some of the limitations in 3D visualizations, such as the effects of occlusion and difficulties interacting with the visualization. The input to the visualization, that is, the 2D projection, is done employing later techniques, capable of high degree of precision in terms of point neighborhoods. Also, the produced landscapes are enriched by visual mappings that support summarization and have being employed successfully in various 2D applications.

The result of this effort is a revisited and improved landscape display for visualizing and exploring a set of data objects. We present a pipeline for the construction of the surface landscapes for multidimensional data, supported by newer projection techniques and visually improved by segmentation algorithms, guiding lines and textures connected to the data, as well as by a 2D cursor for interaction. The result is amenable to user exploration, and reinforces our belief that geographic metaphors can be an effective exploratory environment for visual interpretation data sets, taking advantage of the third dimension while keeping users mental references stable.

The rest of this article is organized as follows. In Section 2, we recall the related work on landscape views and we group the issues surrounding the 2D/3D question. In Section 3, we present our pipeline, together with examples of landscapes for non-hierarchical data. In Section 4, we present our concluding remarks.

2 Related work

The geographic metaphor, where data that is not geographical in nature are represented as a landscape, is recurrent in information visualization. The intention is that the user should benefit from the familiar notion of a map and could then explore data objects guided by neighboring relations intrinsically imposed by geometry and represented by mountains and valleys. Fabrikant et al. (2010) suggests that landscape layouts work not because people understand geomorphological landscapes but because "everyday experience with manipulable tabletop spaces is extended metaphorically to a wide variety of domains of other spatial scales", and then people associate higher with more and bigger with more.

Chalmers (1993) proposed reducing the dimensionality of a space obtained from the weighted term-frequency of a document collection, obtaining a 3D layout to which a horizon is added. The resulting layout resembles a map where documents are shown as icons. He points out advantages of the approach, such as the wide overview provided on the data, the representation continuity, and the preservation of the mental model. A remarkable point of that article is the reference to the book by Lynch (1960), that enumerates elements in a city map which are strongly related to the ability of building a city image and were identified by persons during surveys. Such elements are paths, edges (linear elements that are not paths), districts (medium to large portions of the city), nodes (strategic points which the observer may enter while traversing the city) and landmarks (reference points). Lynch remarks that continuity and contrast should coexist and balance, and that a city view should be stable over time. He also states that such elements are present in models of more general environments, not only in models for cities.

Wise et al. (1995) introduced ThemeScape, that starts from a collection of textual documents, builds a vector representation and then applies a multidimensional reduction to obtain a 2D space. Theme strength of each point is mapped to the third dimension to produce a landscape. The authors highlight that the layout is suitable for revealing interrelationships among documents in either large or small scale. Using a multidimensional reduction to obtain a map is a prevalent procedure, although computationally expensive in the past.

There are many multidimensional reduction techniques in the literature, such as principal component analysis (Jolliffe, 2002), self-organizing maps (Kohonen et al., 2001) force-directed placement (Fruchterman & Reingold, 1991) and multidimensional projections. More recently a large improvement has been achieved on techniques for multidimensional projections resulting in algorithms that are general, fast and precise. This improves on previous algorithms that were either slow, subject to large deviations regarding the original space or restricted to data having specific patterns. Among these techniques we cite ProjClus (Paulovich & Minghim, 2006) and Least-Square Projection (LSP – Paulovich et al., 2006).

Skupin & Buttenfield (1996) proposed building landscapes for newspaper articles applying MDS to a binary vector space of term presence in articles, then mapping the 3rd dimension to the number of terms in each article. They report using a geographic information system for rendering and providing interaction with the display, both in 2D and in 3D.

Davidson et al. (1998) introduced a system (VxInsight) for information visualization that represents data and their relationships as a graph and them produces displays through Laplacian eigenvectors or force-based placement of vertices. A landscape is built mapping density to the 3rd dimension. The display may be augmented by showing edges between objects and by labeling.

Boyack et al. (2002) build a 2D space from scientific articles applying a force-directed scheme to text attributes, and then add a third dimension that reflects density, obtaining a landscape. They also allow adding arrows between data points to indicate citations, augmenting the display with different symbols.

Bischoff et al. (2004) suggest combining ThemeScape with the ThemeRiver (Havre et al., 2002) to analyze a document collection over time. They also superimpose a tree on the landscape to highlight cross-references among documents.

Skupin (2004) builds a multi-level 2D landscape for a set of scientific abstracts in cartography. The lowest zoom level has the documents represented as points. The next zoom level groups points related by the occurrence of specific terms into clusters. Each subsequent zoom level groups clusters in the previous one under more general terms. The highest zoom level depicts large clusters that represent broad terms in the area. Hall & Clough (2013) build a similar visualization.

Jaffe et al. (2006) introduced Tag Maps, that superimpose a word-cloud on a map. Their application used word-clouds built from tags added to photos that were geo-referenced. They report that users were positive regarding Tag Maps for summarizing a map region.

In a comprehensive study on points and landscapes, Tory et al. (2007) compared seven different dot and landscape representations of data: colored and greyscale points; colored and greyscale 2D landscapes; colored, greyscale and uncolored 3D landscapes. Users were evaluated in the task of selecting a display region that contained the most points of a specified target value range, which were most often represented by a single color or grey value. Landscape displays varied in visual complexity (regarding levels) along the experiments. Users were asked to rate the displays with respect to the task they performed. The authors concluded that colored points were the most accurate, fast and highly rated display. Landscapes were from 4 to 10% less accurate and from 1.9 to 4.2 times slower than points. They concluded that 2D was faster than 3D landscapes with no difference in accuracy, and hypothesize that occlusion and the need to rotate the display may have contributed to the difference. They also reported that color displays were faster than greyscale displays, and that uncolored 3D displays were the worst.

In a follow-up experiment, Tory et al. (2009) found that people memorized dot displays more accurately than 2D or 3D landscapes, hypothesizing that the extra features of landscapes may be distracting. 3D landscapes were found to be more accurately memorized than 2D landscapes. They also reported that in most cases the memory accuracy increased with points density.

In the work of Jianu & Laidlaw (2013), a set of genes is represented on a map such that the distance is proportional to the dissimilarity of expression profiles under multiple biological conditions. The interaction with the map is done though zoom and pan. Different levels are represented with a different layout, for instance, point, glyphs and heatmaps.

Similar to the landscape metaphor is the islands metaphor, that includes water and islands to the map, and was used by Pampalk (2001) to visualize music datasets.

Much work in visualization points to the differences among 2D, 3D and combined displays, regarding both the user ability to interact with the display and user's performance on exploratory tasks. There seems to be an undisputed matter the conclusion that 3D displays, or the combination of 3D and 2D displays, are superior when the task is to visualize models of real world objects, scenes or phenomena.

In other contexts, including in information visualization, the use of 3D has received critics that range from mild to acute. The negative characters of 3D layouts that are most frequently pointed out are the occlusion of objects, and the difficulty to maintain a reference with respect to the viewing angle and to the space axes, impairing navigation and orientation. Dealing with such issues demands, for instance, removing layers, using transparency or providing additional views, at the cost of visual stability and context loss, and additional complexity to the user.

As an example, Cockburn & McKenzie (2001) have noticed no improvement of 3D over 2D data mountains (Robertson et al., 1998) in tasks of organizing and retrieving web pages represented as thumbnails. Westerman & Cribbin (2000) concluded that 2D is more effective than 3D for searches in a virtual space built from data, but may be better for browsing and to visualize complex relationships among data.

Risden et al. (2000) compared the performance of specialized users in searching and updating categories of web-page indices both presented as lists and sublists in 2D and as 3D graphs in a hyperbolic display. They observed that 3D improved speed while preserved precision in search tasks, but 2D was more effective for the creation of new categories.

Piringer et al. (2004) analyzed scatter-plots in 2D and in 3D, and remark the difficulty of recognizing point density and depth in 3D views. They propose alternative representations for the points to improve point discrimination, but note that for dense spaces many data objects may be represented by a single pixel, which is a limitation that is difficult to circumvent. Tory et al. (2006) analyzed 2D, 3D and combined displays for estimating relative position, orientation and volume of objects, and conclude that 3D is suitable for approximated navigation and positioning, but are not precise in general for precise positioning and navigation. In his book Mazza (2009) discourages the use of 3D except for real-world visualizations.

Fabrikant et al. (2014) conducted a study with network spatializations and found no improvement in the ability of users to identify similar documents in 3D compared to 2D, hypothesizing that the extra information provided by 3D may not be worth the extra cognitive effort.

Many of the problems encountered with 3D user studies refer to lack of reference to support recognizing the patterns in 3D with the same models as in 2D, and with difficulties in interaction. However, along the last years there were very informative and successful visualizations that employed the third dimension to allow users to interpret additional information or to reduce information loss due to mapping from higher dimensions.

One such example shows the use of 2.5 visualizations to explore biological networks (Fung et al., 2008). While graphs in general are particularly difficult to understand from a 3D representation, a possible 2.5D alternative is to draw parts of the graph in layers, offering thus references for recognition. The article shows a possible rendering for a network with various ‘piled up’ slices, each one with a sector of the graph that is connected to the other slices or planes through graph edges.

Another example of 3D representation shows a mapping from multidimensional scaling algorithms onto 3D space rather than the usual 2D space (Nam & Mueller, 2013). The resulting 3D point cloud is difficult to interpret statically, but is incremented by the possibility of rotation, counting on motion parallax to allow the user to interpret proper clusters in the data. Interaction can be done over any 2D view from the 3D plot. Another work shows an alternative way to interact with 3D clouds, via clustering, which eases selection by users allowing them to drill down a particular cluster formed in the projection (Poco et al., 2012).

Although, for abstract datasets, 2D or point views are more common, the use of the third dimension (or third coordinate plane) can and most likely should be considered to allow additional representational power to the user.

In that sense, landscape views, besides being a common representation for actual

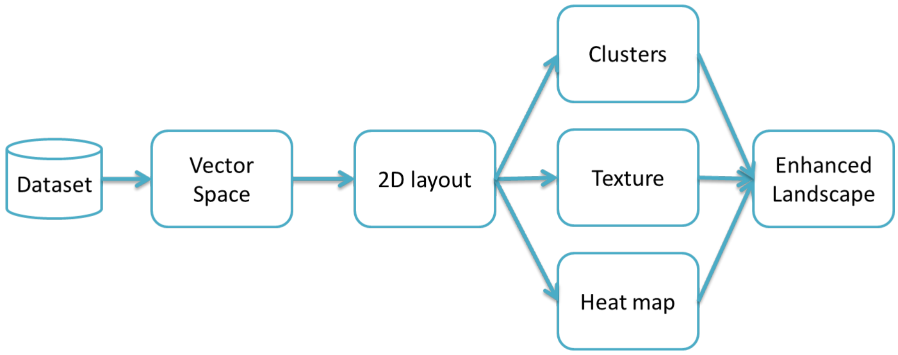


Figure 1: Enhanced landscape construction pipeline.

terrain and geographical applications, can provide support to interpretation of various types of information. Some of the advantages are:

- supports perception of distribution according to a selected concept or variable;
- allows global description of similarity data mapped to 2D planes;
- lends itself to summarization of datasets;
- through artifacts describing regions, levels and distributions, potentially generates additional insights during exploration.

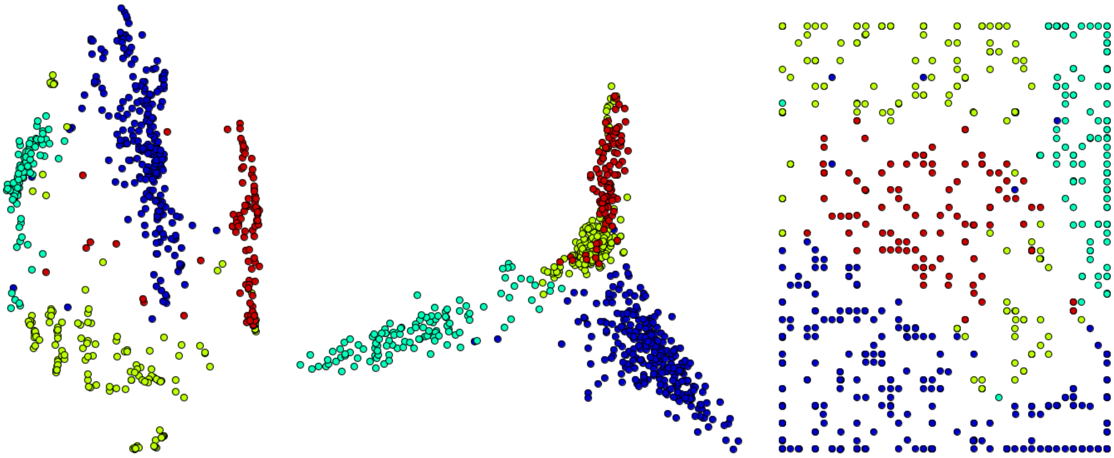
Current 2D mappings based on multidimensional projections have evolved greatly in recent years, and it is our argument that the new capabilities of these recent approaches improve the potential for coherent and informative landscape views. The reason is twofold:

- recent projections are more precise, forming groups of similar objects with less disturbance of neighborhoods, which eases exploration, and
- projections are nowadays also faster, making landscape views an alternative analysis tools for multilevel interaction

In this article we illustrate the pipeline for creating and presenting landscape views that take advantage of the descriptive nature of multidimensional projections and demonstrate the artifacts that support better landscape exploration for applications such as interpretation of text and image collections.

3 Enhanced landscape construction

Starting from a set of data objects, we build a landscape through the pipeline shown in Figure 1. Each data object is processed to obtain a vector, that typically has many dimensions. Then a projection technique is used to obtain a 2D space with point proximity based on similarity, which in turn is clustered and segmented. An additional data dimension is used to generate a landscape for display. If data objects are representatives of image or text data sets, then a word cloud or an image mosaic may be used as landscape texture. We discuss each step of the pipeline next.



(a) LSP (b) PCA (c) SOM

Figure 2: Projections of CBR+. Color indicates pre-classification: CBR is dark-blue, ILP if light-blue, IR is yellow and SON is red.

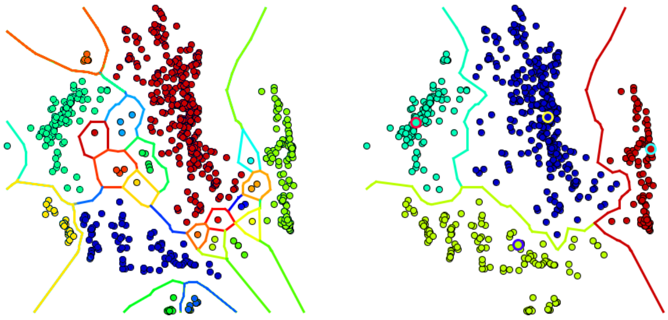
Data and Vector Space Obtaining a vector from each data object typically involves extracting features and evaluating some measures. For text, term frequency-inverse document frequency (TF-IDF) is often used (Salton & Buckley, 1988). For images, descriptors like SIFT (Li & Wang, 2003) and BIC (Stehling et al., 2002), as well as many others, may be applied to obtain descriptive vectors for objects. The resulting vector space is usually a high-dimensional space.

For instance, a dataset with extracts of scientific articles, containing title, authors, abstract and references, from journals devoted to Case-based Reasoning (CBR), Inductive Logic Programming (ILP), Sonification (SON) and Information Retrieval (IR) was first processed for stop-words removal, stemming, Luhn's cut of terms with highest and lowest frequencies (Luhn, 1958) and then TF-IDF was evaluated. The resulting vector-space has 675 vectors with $1{,}423$ dimensions. We will refer to this dataset as CBR+. The class of each article reflects origin of the document, then, in terms of subjects, some articles may not be in the most suitable class or should belong to more than one class. It is often the case in practice that pre-classification, if one exists, should not be taken strictly.

2D layout - mapping higher dimensions to visual 2D spaces The next step is to project the vector-space, obtaining a 2D space. This is the role of a multidimensional projection, which strives to maintain certain properties of the original space in the projected space. For instance, Figure 2 shows different projections for the CBR+ dataset. The depicted layouts were produced by LSP, Principal Component Analysis (PCA – Jolliffe, 2002) and Self-Organizing Maps (SOM – Kohonen et al., 2001).

Clustering - finding suitable groups of objects in projections Clustering is the next step in the process. The goal is to define groups that allow partitioning the space into regions that will enhance data organization for exploration and navigation in data space.

Although any clustering algorithm may be used, a clustering algorithm that favors a smooth partitioning of the space into regions is more suitable. We propose two algorithms based on a Voronoi diagram. One defines clusters automatically and the other takes user guidance for defining the clusters. Both are very efficient computationally because a Voronoi diagram may be built in $O(n \lg n)$ time and the resulting



(a) Voronoi clustering. (b) Seed clustering.

Figure 3: Voronoi and Seed clustering of CBR+ after LSP (from Figure 2a. The Voronoi clustering distance parameter value is 5% of the largest distance between any two points. In the seed clustering display, seeds are highlighted. Colors indicate formed clusters.

structure is a planar map.

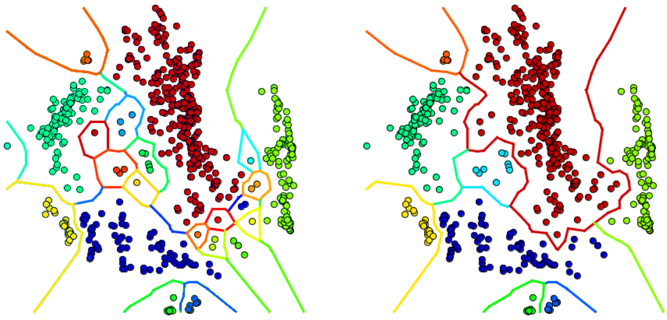
The first algorithm, called Voronoi Clustering, is greedy and distance-based. The algorithm starts with a Voronoi diagram where each point is a site (thus each diagram's region has exactly one data point). The algorithm visits every point p in arbitrary order and considers its neighbors. If the distance between p and a neighbor q is smaller than an input threshold then the regions of p and q are united.

The second algorithm, called Projection Seed Clustering, relies on the user to provide cluster seeds, visually selecting points on the projection display. Each cluster seed will give rise to a cluster. Suppose a graph G where each site is a vertex and such that there is a weighted edge between two vertices if their region in the Voronoi diagram is adjacent. Edge weights are distances. For each non-seed point p, evaluate the shortest-path from p to each seed s, and record the edge e_s in the shortest-path having the largest weight. Then add p to the region of the seed whose path from p contains the edge with smallest weight among those recorded for the shortest-paths, breaking ties arbitrarily. Although this algorithm resembles a shortest-path evaluation in a graph, the graph must not be built explicitly; using the Voronoi diagram and a queue leads to a very efficient implementation.

Figure 3 shows the Voronoi and seed clustering for CBR+ applied after LSP.

When the input is a pre-classified dataset, it may be an option to skip clustering altogether. After projection, the 2D space may be segmented to produce a partition of the display, as those obtained with the clustering algorithms described above. The following algorithm will produce a segmentation of the 2D space from labeled (colored) points, with the premise that distance in the layout takes precedence over class information. Starting with a Voronoi diagram where each point is a site, unite adjacent regions whose site belong to the same class. Then, while either there are regions with too few points or there is a large number of regions, unite the smallest region with the closest one. An example output is shown in Figure 4.

There are other scenarios for using the above segmentation algorithm. One of them is to reflect, in 2D space, clusters that approximate a cluster in the original space. Points are clustered in original space, which may give a precision gain, projected in 2D, and then subject to this segmentation algorithm, taking each cluster as a class, to obtain a more regular partitioning of the space into regions. Another scenario is using this segmentation to reduce the number of clusters obtained with Voronoi clustering or any other clustering algorithm.



(a) Before segmentation. (b) After segmentation.

Figure 4: The application of segmentation on a previously clustered space. The number of clusters was reduced by grouping small clusters to the closer ones.

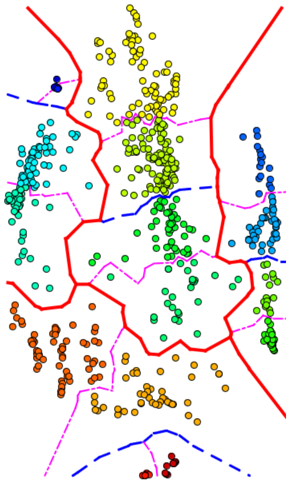
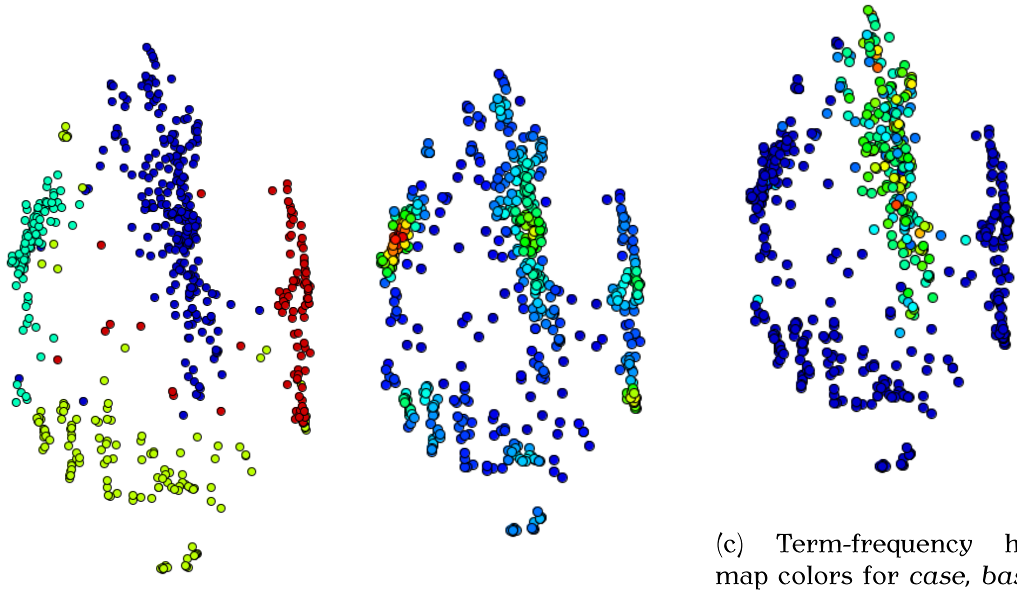


Figure 5: A region partitioned into three levels of clustering. Clusters are divided by a thick continuous curve in the first level, by a blue discontinuous curve in the second level and by a thin discontinuous magenta curve in the third level.

The clustering algorithms described above may be applied recursively to refine the layout in as many levels as wanted, limited by the number of levels or the number of points in each region. For instance, Figure 5 shows a diagram partitioned in three levels. Such multi-level clustering naturally leads to multi-level presentations, supporting the generation of landscape layouts in larger scale.

Landscaping The layout may be enriched by mapping point color or attribute to height, thus generating a surface whose shape is indicative of data behaviour. We have considered two different heat maps that help the exploration of the layout. One is generally applicable: a density heat map. The other, for text, is a term-frequency heat map. Both are constructed evaluating a scalar for each point and mapping such scalars onto a color scale. For instance, in Figure 6 the CBR+ dataset is shown under different heat maps. The scalar for density was evaluated using the Epanechnikov kernel function (Silverman, 1986). The scalar for term-frequency was evaluated as the number of occurrences of one or more words (user input) in a document divided by the number of words in the document.

The landscape is built from the 2D space mapping one attribute to the third dimension. For instance, Figure 7 shows a view of a landscape built for CBR+ with



(a) Class colors. (b) Density heat map colors. (c) Term-frequency heat map colors for *case*, *based* and *reasoning*.

Figure 6: Heat maps for CBR+.

density mapped to height. The cursor is positioned on a point, whose neighbors are highlighted by connecting lines and which label is shown. This type of cursor has been known as spider cursor (Minghim et al., 2005). Other attributes may be used to define height as well, including search terms for texts. This layout allows a global analysis relating the dataset to an attribute (e.g. group density or term) as well as individual analysis of both groups and points.

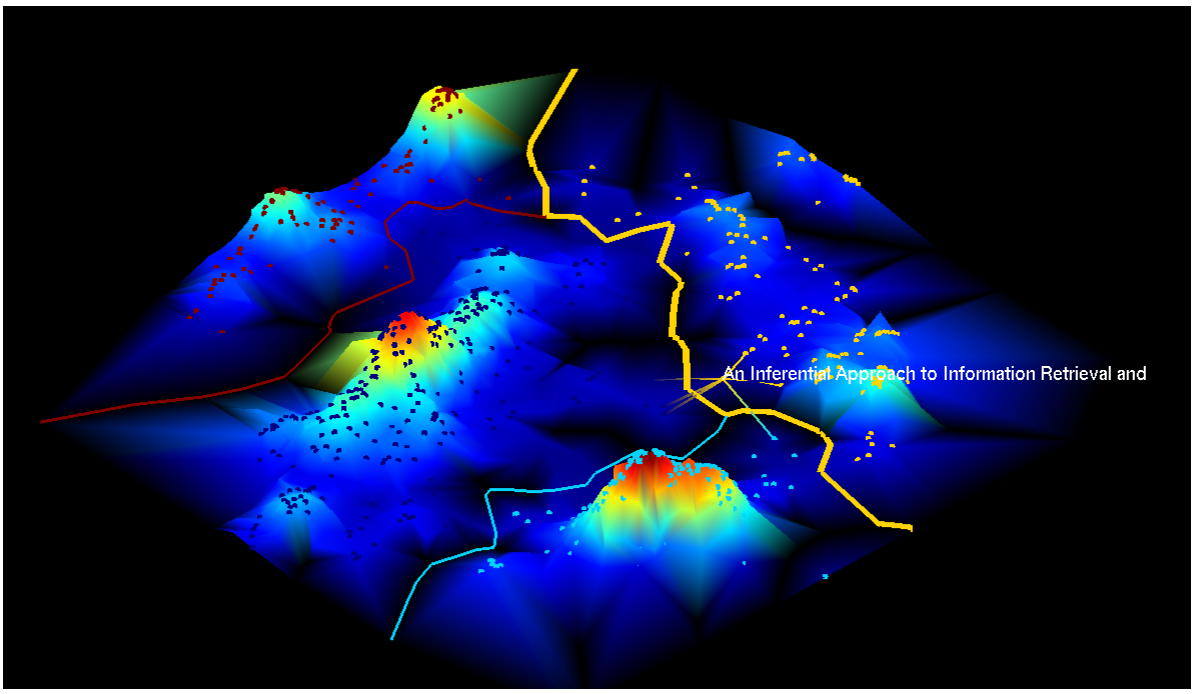


Figure 7: Landscape for CBR+, with density mapped to height.

We also propose combining the landscape with a texture that summarizes regions and adds more information to the layout, enhancing navigation and exploration. For text datasets, word clouds will provide a perfect texture for exploration. Figure 8 shows a view of the CBR+ density landscape enhanced by a word cloud built with the algorithm proposed by Paulovich et al. (2012), which generates word clouds in planar regions. Rotating the view allows alternating between a clearer view of the landscape slopes and valleys and a clearer view of words in each region. The height factor can also be removed when the association with the mapped attribute is no longer of interest.

The bordering lines of regions defined by clusters and segmentation also improve navigation and help preserving the mental model. Borders may enhance the perception of regions of interest and help preserving reference points as the layout is rotated and zoomed, alleviating the effects of occlusion and flatness imposed by angled perspectives. Groups can be selected at once via the boundaries of the regions detected in the current segmentation. In the figures those boundaries are shown as linear curves.

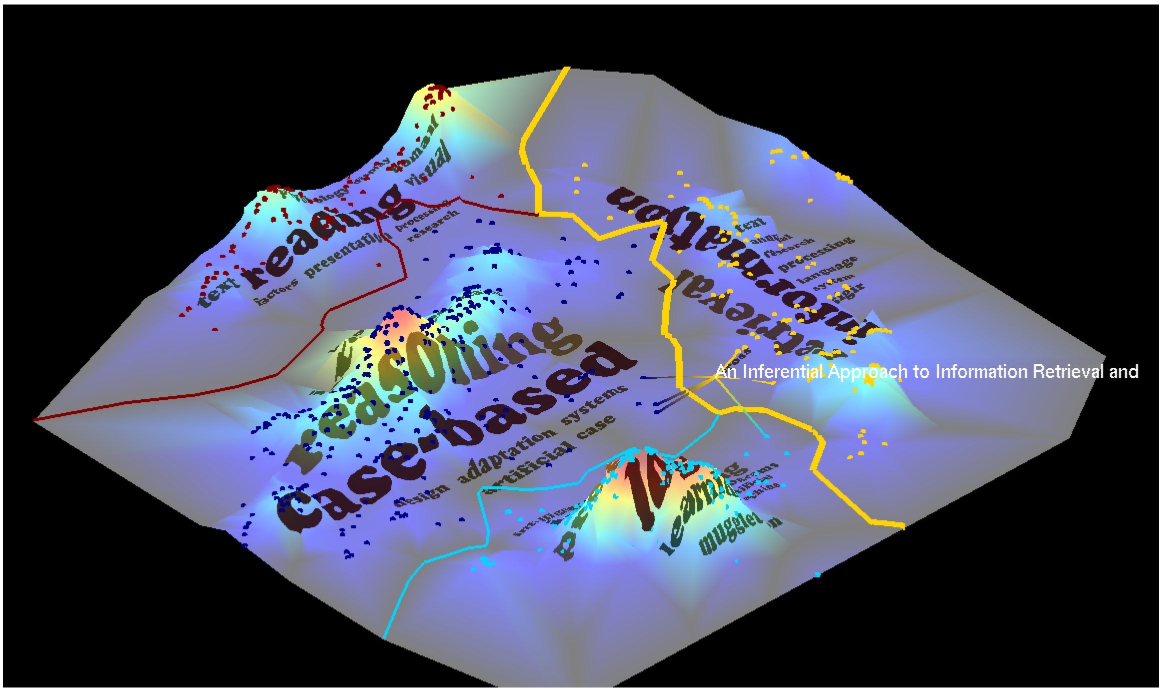


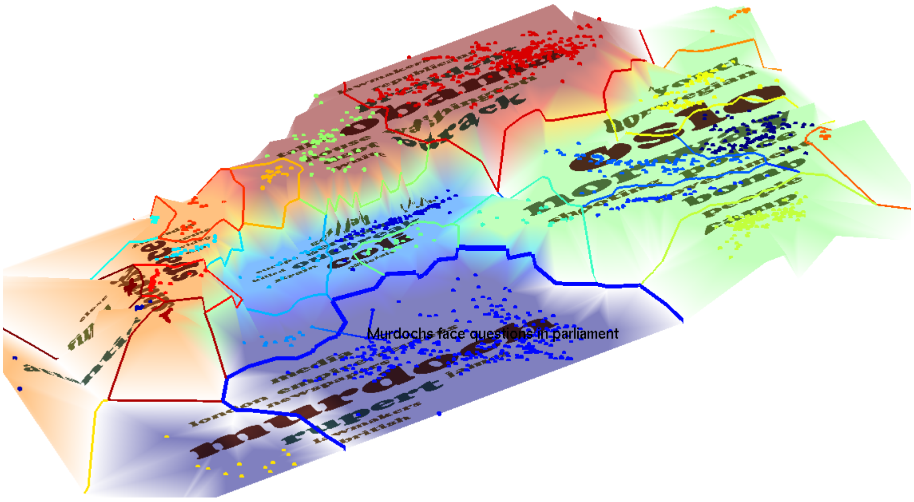
Figure 8: Landscape with word cloud texture and clusters for CBR+, with density mapped to height.

Another example appears in Figure 9, that shows a landscape for a news dataset consisting of 1,771 RSS news feeds collected from BBC, CNN, Reuters and Associated Press during June and July of 2011. After processing, each document was turned into a vector of 1,084 dimensions.

A natural texture for image datasets is an image mosaic, such as those proposed by Tan et al. (2012). Figure 10 illustrates the concept for images in a sub-set of the Corel dataset that has 1,000 images divided in 10 classes (African tribes, beaches, buildings, buses, dinosaurs, elephants, flowers horses, mountains and food), from which 150 dimensions were extracted through SIFT descriptors.

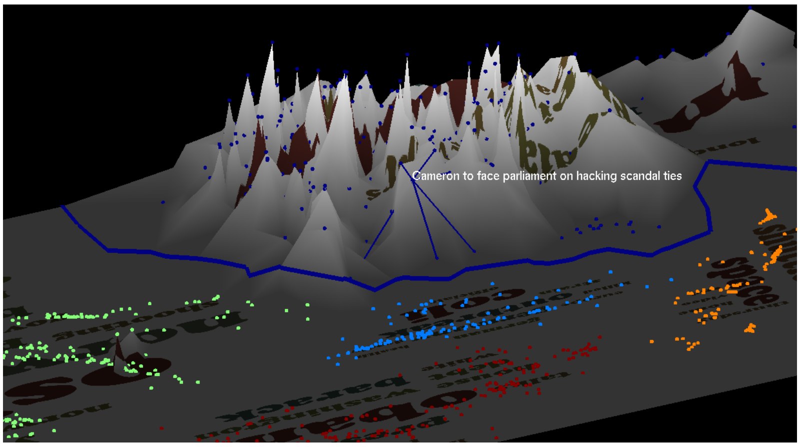
A summary may also be provided for a set of videos resulting also in a mosaic of images extracted from the video by summarizing algorithms. Music or other sound recordings may be summarized by a word cloud of tags, with added aural summaries played when a cursor is positioned on regions or points.

In general, a dataset may be summarized in many ways. Numeric attributes may be



(a) News landscape with word clouds, and density as height.

(b) Flatter version of the news landscape.



(c) News landscape with search terms frequencies (murdoch and hacking) mapped to height.

Figure 9: Landscapes from RSS news feeds.

mapped on the landscape themselves or displayed as dispersion curves or histograms, for instance. Categorical attributes may be displayed as tag clouds as well, for instance.

Naturally, on the top of the landscape displays that we propose here, a series of symbols and glyphs may be added to enhance some characters of each dataset. While the effectiveness of adding more elements to the display must be considered for each dataset, particular applications may benefit from additional information on the display.

Finally, the landscape display, coupled with summarization and segmentation algorithms, may open the possibility for multi-level presentations. For instance, summarization can occur in various levels. While at the top level a general view, as well as the large parts of the dataset, are displayed, the user may choose a region to focus his or her exploration. Figure 11 shows a top view of the news dataset presented before. In that figure, the top level is presented and subjects of the various regions are displayed via a co-occurrence based topic extraction algorithm. A group is then selected, and for that group, the display is repeated, with an added level of detail given by new clustering followed by segmentation. The strategy is prone to be applied for multi-level point-based and surface-based displays.

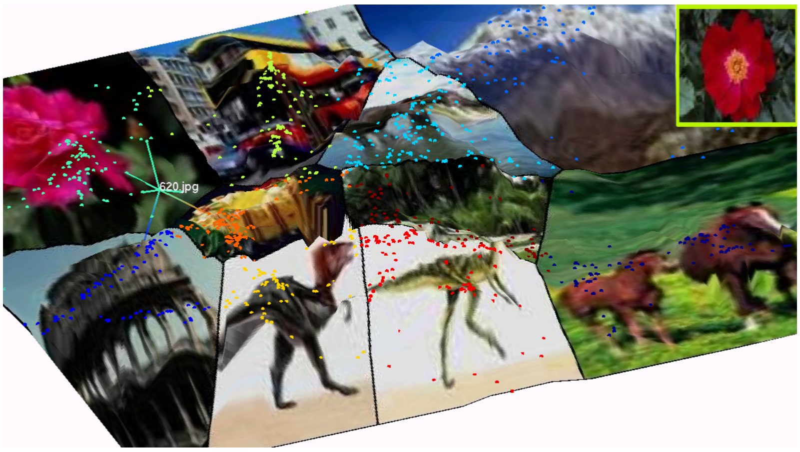
4 Conclusion

Landscapes for information visualization have been encouraged by many authors, although without support from experiments with user subjects. We believe that the latest projection techniques that have been developed recently renew the strength of mapping data through the concept of landscapes, as they are both fast and precise enough to enable real-time construction of layouts recursively and in a very large scale. Visual guidance, as those we introduced in this article, namely the division of the landscape in regions and the use of texture, also help in avoiding disorientation in 3D and increases contextual overviews. The view lends itself to multi-level visualizations, allowing for future development of visualization in larger scale.

Acknowledgements

We acknowledge the financial support of FAPESP and CNPq Brazilian research financial agencies for the development of this work.

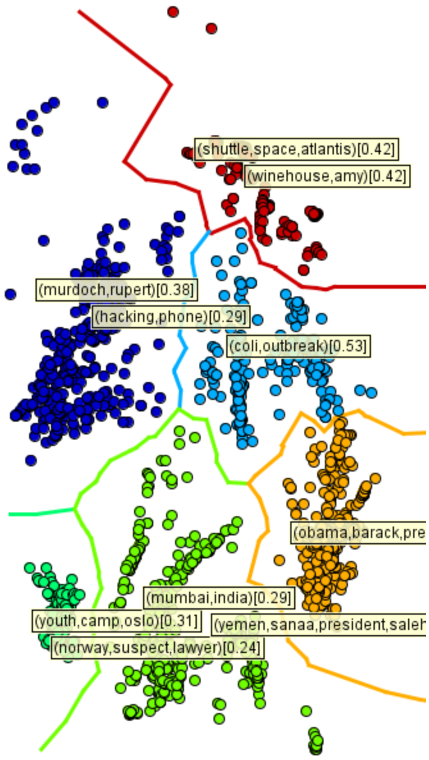
(a) Mosaic over projection. Central picture of the cursor is shown.



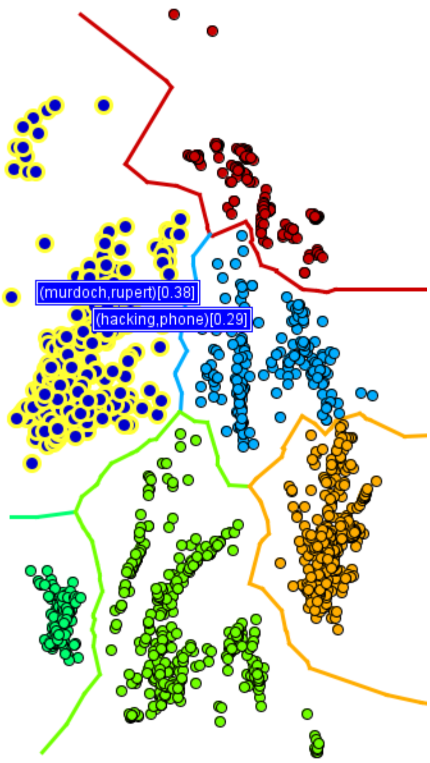
(b) Mosaic with density landscaping. Central picture of the cursor is shown.

(c) Mosaic with display of all pictures touched by the spider cursor.

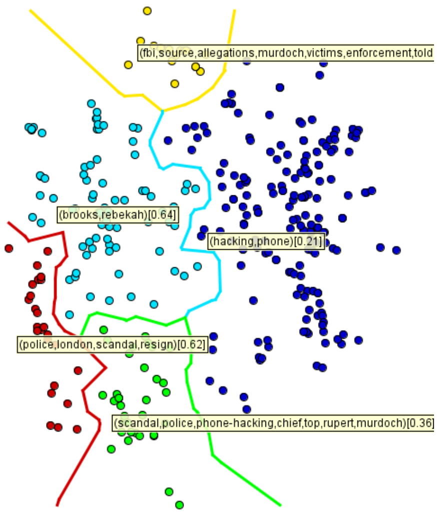
Figure 10: Landscape with mosaic texture and clusters for the Corel dataset, a 10-class image dataset. Point colors indicate classes.



(a) News map with groups of news subjects, identified by topics.



(b) Selection of a group of news feeds.



(c) New level of grouping in the selected set, after clustering and segmentation.

Figure 11: Drilling down on the map levels.

References

Bischoff, U., Diakopoulos, N., Loesch, F. & Zhou, Y. (2004), ThemeExplorer: A tool for understanding the history of the field of information visualization, *in* 'Proc. InfoVisFun'.

Boyack, K. W., Wylie, B. N. & Davidson, G. S. (2002), 'Domain visualization using Vx-Insight for science and technology management', *Journal of the American Society for Information Science and Technology* **53**(9), 764–774.

Chalmers, M. (1993), Using a landscape metaphor to represent a corpus of documents, *in* 'Proc. European Conference on Spatial Information Theory', pp. 377–390.

Cockburn, A. & McKenzie, B. (2001), 3D or not 3D?: Evaluating the effect of the third dimension in a document management system, *in* 'Proc. SIGCHI Conference on Human Factors in Computing Systems', pp. 434–441.

Davidson, G. S., Hendrickson, B., Johnson, D. K., Meyers, C. E. & Wylie, B. N. (1998), 'Knowledge mining with vxinsight: Discovery through interaction', *Journal of Intelligent Information Systems* **11**, 259–285.

Fabrikant, S. I., Maggi, S. & Montello, D. R. (2014), 3D network spatialization: Does it add depth to 2D representations of semantic proximity?, *in* 'Proc. GIScience', pp. 34–47.

Fabrikant, S. I., Montello, D. R. & Mark, D. M. (2010), 'The natural landscape metaphor in information visualization: The role of commonsense geomorphology', *Journal of the American Society for Information Science and Technology* **61**(2), 253–270.

Fruchterman, T. M. J. & Reingold, E. M. (1991), 'Graph drawing by force-directed placement', *Software - Practice and Experience* **21**(11), 1129–1164.

Fung, D. C., Hong, S.-H., Koschutzki, D., Schreiber, F. & Xu, K. (2008), '2.5d visualisation of overlapping biological networks', *Journal of Integrative Bioinformatics*.

Hall, M. & Clough, P. (2013), Exploring large digital library collections using a map-based visualisation, *in* 'Proc. International Conference on Theory and Practice of Digital Libraries', pp. 216–227.

Havre, S., Hetzler, E., Whitney, P. & Nowell, L. (2002), 'Themeriver: visualizing thematic changes in large document collections', *IEEE Transactions on Visualization and Computer Graphics* **8**(1), 9–20.

Jaffe, A., Naaman, M., Tassa, T. & Davis, M. (2006), Generating summaries and visualization for large collections of geo-referenced photographs, *in* 'Proc. ACM International Workshop on Multimedia Information Retrieval', pp. 89–98.

Jianu, R. & Laidlaw, D. H. (2013), 'What google maps can do for biomedical data dissemination: examples and a design study', *BMC Research Notes* **6**(1), 1–14.

Jolliffe, I. T. (2002), *Principal Component Analysis*, 2 ed., Springer-Verlag.

Kohonen, T., Schroeder, M. R. & Huang, T. S., eds (2001), *Self-Organizing Maps*, 3rd ed., Springer-Verlag.

Li, J. & Wang, J. Z. (2003), 'Automatic linguistic indexing of pictures by a statistical modelin approach', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **25**, 1075–1088.

Luhn, H. P. (1958), 'The automatic creation of literature abstracts', *IBM Journal of Research and Development* **2**(2), 159–165.

Lynch, K. (1960), *The Image of the City*, MIT Press.

Mazza, R. (2009), *Introduction to Information Visualization*, Springer-Verlag.

Minghim, R., Levkowitz, H., Nonato, L. G., Watanabe, L., Salvador, V., Lopes, H., Pesco, S. & Tavares, G. (2005), Spider cursor: A simple verstile interaction tool for data visualization and exploration, *in* 'Proceedings of the 3rd International Conference on Computer Graphics and Interaction Techniques in Australasia and Southeast Asia - GRAPHITE', ACM Press, Dunedin, New Zeland, pp. 307–314.

Nam, J. E. & Mueller, K. (2013), 'Tripadvisor N-D: A tourism-inspired high-dimensional space exploration framework with overview and detail', *IEEE Transactions on Visualization And Computer Graphics* **19**(2), 291–305.

Pampalk, E. (2001), Islands of Music Analysis, Organization, and Visualization of Music Archives, PhD thesis, Technischen Universität Wien.

Paulovich, F. V. & Minghim, R. (2006), Text map explorer: a tool to create and explore document maps, *in* 'Proc. IEEE Symposium on Information Visualization', pp. 245–251.

Paulovich, F. V., Nonato, L. G., Minghim, R. & Levkowitz, H. (2006), Visual mapping of text collections through a fast high precision projection technique, *in* 'Proc. IEEE Symposium on Information Visualization', pp. 282–290.

Paulovich, F. V., Telles, G. P., Toledo, F. M. B., Minghim, R. & Nonato, L. G. (2012), 'Semantic wordification of document collections', *Computer Graphics Forum* **31**, 1145–1153.

Piringer, H., Kosara, R. & Hauser, H. (2004), Interactive focus+context visualization with linked 2D/3D scatterplots, *in* 'Proc. International Conference on Coordinated and Multiple Views in Exploratory Visualization', pp. 49–60.

Poco, J., Eler, D. M., Paulovich, F. V. & Minghim, R. (2012), 'Employing 2D projections for fast visual exploration of large fiber tracking data', *Computer Graphics Forum* **31**(3), 1075–1084.

Risden, K., Czerwinski, M. P., Munzner, T. & Cook, D. B. (2000), 'An initial examination of ease of use for 2D and 3D information visualizations of web content', *International Journal of Human-Computer Studies* **53**(5), 695–714.

Robertson, G., Czerwinski, M., Larson, K., Robbins, D. C., Thiel, D. & van Dantzich, M. (1998), Data mountain: Using spatial memory for document management, *in* 'Proc. UIST'.

Salton, G. & Buckley, C. (1988), 'Term-weighting approaches in automatic text retrieval', *Information Processing & Management* **24**(5), 513–523.

Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, Chapman & Hall.

Skupin, A. (2004), 'The world of geography: Visualizing a knowledge domain with cartographic means', *PNAS* **101**(1), 5274–5278.

Skupin, A. & Buttenfield, B. P. (1996), Spatial metaphors for visualizing very large data archives, *in* 'Proc. GIS-LIS International Conference', pp. 607–617.

Stehling, R. O., Nascimento, M. A. & Falcão, A. X. (2002), A compact and efficient image retrieval approach based on border/interior pixel classification, *in* 'Proc. International Conference on Information and Knowledge Management', pp. 102–109.

Tan, L., Song, Y., Liu, S. & Xie, L. (2012), 'Imagehive: Interactive content-aware image summarization', *IEEE Computer Graphics and Applications* **32**(1), 46–55.

Tory, M., Kirkpatrick, A. E., Atkins, M. S. & Moller, T. (2006), 'Visualization task performance with 2D, 3D, and combination displays', *IEEE Transactions on Visualization and Computer Graphics* **12**(1), 2–13.

Tory, M., Sprague, D. W., Wu, F., So, W. Y. & Munzner, T. (2007), 'Spatialization design: Comparing points and landscapes', *IEEE Transactions on Visualization and Computer Graphics* **13**(6), 1262–1269.

Tory, M., Swindells, C. & Dreezer, R. (2009), 'Comparing dot and landscape spatializations for visual memory differences', *IEEE Transactions on Visualization and Computer Graphics* **15**(6), 1033–1040.

Westerman, S. J. & Cribbin, T. (2000), 'Mapping semantic information in virtual space: Dimensions, variance and individual differences', *International Journal of Human-Computer Studies* **53**(5), 765–787.

Wise, J. A., Thomas, J. J., Pennock, K., Lantrip, D., Pottier, M., Schur, A. & Crow, V. (1995), Visualizing the non-visual: Spatial analysis and interaction with information from text documents, *in* 'Proc. IEEE Symposium Information Visualization', pp. 51–58.