Geo-Scape, a Granularity Depended Spatialization Tool for Visualizing Multidimensional Data Sets

Kontaxaki Sofia, Kokla Margarita, Kavouras Marinos

School of Rural and Surveying Engineering, National Technical University of Athens, 9 H. Polytechniou Str., Zografos Campus, 15780 Athens, Greece

© Wuhan University and Springer-Verlag Berlin Heidelberg 2010

Abstract

Recently, the expertise accumulated in the field of geovisualization has found application in the visualization of abstract multidimensional data, on the basis of methods called spatialization methods. Spatialization methods aim at visualizing multidimensional data into low-dimensional representational spaces by making use of spatial metaphors and applying dimension reduction techniques. Spatial metaphors are able to provide a metaphoric framework for the visualization of information at different levels of granularity. The present paper makes an investigation on how the issue of granularity is handled in the context of representative examples of spatialization methods. Furthermore, this paper introduces the prototyping tool Geo-Scape, which provides an interactive spatialization environment for representing and exploring multidimensional data at different levels of granularity, by making use of a kernel density estimation technique and on the landscape “smoothness” metaphor. A demonstration scenario is presented next to show how Geo-Scape helps to discover knowledge into a large set of data, by grouping them into meaningful clusters on the basis of a similarity measure and organizing them at different levels of granularity.

Keywords

multidimensional data; spatial metaphors; spatialization; graphical interface; kernel density estimation

Introduction

Geovisualization is commonly used for the purpose of visualizing and interacting with geospatial data, namely the data that have either spatial properties or are georeferenced. However, recently, the expertise accumulated in the field of geovisualization, and particularly of cartography,\(^1\) has found application in the visualization of abstract multidimensional data, on the basis of methods called spatialization methods. Spatialization methods aim at visualizing multidimensional data into low-dimensional representational spaces by making use of spatial metaphors.

Multidimensional data are large sets of data observations on two or more variables, inherent in a broad range of applications such as bioinformatics, statistics, commercial applications, geographic information systems, and so on. In order to be represented into a two- or three-dimensional representational
space, multidimensional data have to be projected according to techniques that reduce the number of dimensions. Spatialization methods utilize such techniques to perform the dimension reduction.

Another difficulty that the spatialization methods have to face is the representation of information at different levels of granularity. Granularity usually refers to the different scales or levels of detail information can be represented at. This paper intends to discuss the way spatialization methods deal with granularity and goes a step further by introducing the prototyping tool Geo-Scape, which provides an interactive environment for representing and exploring multidimensional data at different levels of granularity by making use of the landscape smoothness metaphor.

The rest of the paper is organized as follows: Section 1 performs an investigation of how the existing spatialization methods handle the issue of granularity and gives some representative examples of related work. Section 2 introduces the functionality of Geo-Scape and demonstrates how the exploration of information at different levels of granularity is supported by the provided operations. Section 3 presents a demonstration scenario, which describes how information is visualized through the different views of Geo-Scape. Finally, in Section 4, a discussion is made to expose some negative and positive aspects of the proposed spatialization environment.

1 Granularity in spatialization-related work

Granularity refers to the different scales or levels of detail, information can be represented at. In the context of geovisualizations, granularity changes usually happen as a result of generalization and specialization operations application. The generalization operation reduces the number of visualized features by clustering and replacing them by fewer and less detailed one, on the basis of distance and semantic criterions. The specialization operation consists of exactly the inverse procedure. It involves the splitting of features into a more detailed representation of space.

In such a context, hierarchical clustering is a procedure that appears to be very supportive for the execution of the generalization and specialization operations, since it performs the clustering of the geovisualization features on the basis of the distance assessed between them. Hierarchical clustering leads to the creation of a hierarchical structure called dendrogram. Actually, the dendrogram constitutes an ideal basis for the grouping/splitting of features, inasmuch as each of its levels contains the features to be represented at a specific scale, or granularity level.

A dendrogram defines a hierarchy of clusters beginning from an all-inclusive cluster at the top. Clusters group the objects of a representation on the basis of a distance or similarity measure. The higher a cluster is located in the dendrogram, the more populated it is. However, dendrograms may grow to be so overloaded that the leaf level becomes overcrowded while a great amount of space remains empty and unexploited. An example of an overloaded dendrogram is pictured in Fig.1. Hence, as a method of portraying information at different granularity levels, dendrograms present several drawbacks, which increase with size and may cause problems of readability.

![Fig.1 Example of an overcrowded dendrogram created during hierarchical clustering](image)

The portraying of information at different granularity levels is also of concern of special geovisualization methods called spatializations. In fact, spatialization methods have to confront both the problem of arranging a large amount of multidimensional information into a relatively restricted space and the challenge to provide a metaphoric framework for the visualization of information at different levels of granularity.

Kuhn and Blumenthal define spatialization as a partial mapping between information and physical
space through spatial metaphors. In addition, it is argued that the use of spatial metaphors usually entail the establishment of a hierarchy of sub- and super-metaphors. For example, the concept building may be considered as a sub-metaphor for the super-metaphor city, and city may be a sub-metaphor for the super-metaphor country. In this way, a whole hierarchical structure of metaphors can be used for supporting mappings from data space to spatialization space and by extent, for representing information at different granularity levels.

The metaphor city is extensively examined in Ref.[5]. It is argued that this spatial metaphor, on the one hand, provides hierarchical and dynamic structures that scale relatively well, and, on the other hand, supports the use of sub-metaphors that define clear spatial boundaries (e.g., building, room) and are able to model relationships of containment. These properties of the city metaphor are exploited for the visualization of a large video corpus along multiple dimensions,[6] which are called perspectives. The city metaphor is presented as being a highly memorable metaphor, and so varied and detailed that it is able to convey many conceptual dimensions. Furthermore, another and recent application, the CodeCity tool,[7] also follows the city metaphor for visualizing large-scale software in a 3D representational space. In this case, the sub-metaphors districts and buildings are used to represent respectively the software packages and the specific classes implemented by these packages.

Another example of spatial metaphor is that of an information landscape. Information landscapes make the representational space easy to conceive since it resembles the familiar geographic landscape.[8] “Scapes as found in nature” constitute, according to Benking and Judge,[9] optimal metaphors for modeling, representing, and exploring information at different scales, that is, at different levels of granularity.

The several spatialization environments, which are found in the literature using the landscape metaphor, usually focus on the processing of textual data (abstract, articles, books, web-pages content, etc). Indicative examples are the SPIRE project[10] and the VxInsight application.[11] The SPIRE project performs the spatialization of a corpus of documents but does not provide any functionality either to zoom or explore information at different levels of granularity, or to understand the implicit hierarchical structure of data. The VxInsight application also visualizes documents, patents, or even genomic data, as a landscape. Furthermore, it does allow operations like zooming, to get a more detailed view of a certain part of the landscape.

Another example of application using a metaphor related to the landscape is found in Ref.[12]. More specifically, the use of the island metaphor is proposed for the iterative spatial organization of images initially retrieved from an internet search or from a collection of photos, into a 2D representational space. It can be roughly said that the islands, called visual islands, correspond to the clusters of images that are similar on the basis of some specific set of image features like time, color, texture, etc. Finally, a new graphical representation model based on the landscape metaphor is proposed for the visualization of software projects,[13] and particularly for the controlling and managing of software development processes.[13]

Skupin and Buttenfield[14] describe spatialization as a projection of features belonging to a high-dimensional information space, into a low-dimensional representational space. This second definition implies that dimension reduction techniques have to be used in order to overcome the limitations imposed by the lowdimensional representational space. Self-Organizing Maps (SOM),[16] Principal Component Analysis (PCA)[17] and MDS[18] are well-known projection methods that accomplish the dimension reduction, but their appropriateness for representing information at different granularity levels has yet to be discussed.

Kohonen[16] describes SOM as “a non-linear, ordered, smooth mapping of high-dimensional input data manifolds into the elements of a regular, low dimensional array.” SOM produces similarity graphs of input data, which maintain topological relationships and can be used for many tasks, such as pattern recognition or clustering analysis. A particular visualization of the SOM, called U-matrix (Unified distance matrix),[19] performs clustering analysis by visualizing distance between data using a variation of color hue. The regions of bright color correspond to
data clusters, while the dark regions correspond to the limits of these clusters. Another visualization of SOM is that of a configuration of point-features consisting of a set of distinct objects projected into a two- or three-dimensional space, each represented at a unique position with a specific color and shape. Neighborhood relations can be outlined by lines.[20]

Several applications have exploited the ability of SOM to represent information at different granularity levels. For example, WEBSOM,[21] an application of SOM for the spatialization of large collections of text documents, incorporates multiple scale levels and provides an interactive facility for users to explore the document space. Skupin[22] showed how the resolution of a SOM influences the level of detail of a spatialization. He argued that the finer the resolution, the more detailed the visual representation produced by a SOM, and vice-versa. Furthermore, Skupin[23] has examined how a set of conference abstracts can be represented in a scale-dependent spatialization performed with SOM.

As far as PCA and MDS are concerned, they both preserve the discrete nature of the data being processed, and result in visualizations consisting of point configurations in low-dimensional representational spaces. Specifically, PCA is a well-known statistical method used for information redundancy reduction. Data are projected into a space in which the axes are defined by the eigenvectors of the correlation matrix. PCA manages to convert the initial set of data characteristics into a new set composed of fewer characteristics and generated from a linear transformation of the initial ones. However, PCA is considered to have poor interpretability.[24] MDS uses distance as a measure of similarity. A dissimilarity matrix is created from the semantic comparison of data based on that measure. In the next step, data are visualized as objects represented with labeled points in a low-dimensional information space. The shorter the distance between points, the more similar are the corresponding data.

Nevertheless, neither PCA nor MDS can represent information at different levels of granularity without making use of additional visual variables beyond position. For example, Skupin and Buttenfield[14] created a spatialization of the content of texts based on MDS and the map metaphor. Texts were projected into a two-dimensional space as point features, and at the same time hierarchically clustered into a dendrogram. Each time the currently visualized hierarchical level of the dendrogram changed, the size of the cartographic symbols was readjusted while the map features were merged or split.

Hence, point configurations resulting from dimension reduction techniques, need further processing to model and represent information at different granularity levels. In the next section, a new spatialization environment is introduced to fulfill this need. This environment is provided by Geo-Scape.

2 Geo-Scape

Geo-Scape is a prototyping tool developed by the authors in the context of the OntoGeo Group (http://ontogeo.ntua.gr/) activity, which helps to discover knowledge into a large set of multidimensional data, by grouping them into meaningful clusters on the basis of a similarity measure and organizing them at different levels of granularity.

Geo-Scape is based on a spatialization method[15] that performs kernel density estimation over point configurations resulting from dimension reduction techniques. Specifically, this tool produces three-dimensional surface visualizations according to the landscape metaphor and present information at different levels of granularity by means of an adjustable landscape smoothness. Although there are several applications found in the literature that use the landscape metaphor, as it was presented in Section 1, Geo-Scape differs on the following points:

1. Geo-Scape does not take as input textual data like documents, articles, hypertexts, etc. It focuses on the processing of any kind of multidimensional vector data, as far as they are organized into arrays where each row corresponds to a discrete data item and each column to a different observation.

2. Unlike the other environments using the landscape metaphor, Geo-Scape also makes use of the smoothness metaphor, to change the aspect of the landscape as the granularity level changes and helps to reveal the results of the hierarchical clustering of the data. Hence, the landscape can portray each data
item individually, like a set of “needle-like” mountains, but can also become smoother and smoother, as the clusters gradually merge into fewer ones.

(3) Geo-Scape performs the kernel density estimation based on a triangular function, which helps to make clear the boundaries between the mountains of the landscape. This function will be described next.

(4) Besides the three-dimensional landscape view and the scatter plot view where the data appear as discrete point features, Geo-Scape provides a global view of the entire hierarchy of clusters created among the data, the dendrogram view, which highlights the current level of granularity of the landscape.

(5) Geo-Scape also provides a dynamic labeling strategy and lets the user select one or more representative variables between the several that describe the observations.

The methodological approach and the functionality of Geo-Scape are described in detail next.

2.1 Spatialization method

The spatialization method implemented in Geo-Scape starts with the application of a dimension reduction technique to project the multidimensional data into a two-dimensional space. This version of Geo-Scape currently applies a classical multidimensional scaling method, which assesses a matrix of dissimilarities between pairs of data items and arranges them into the target space to minimize a loss function called stress. In the next step, the method uses kernel density estimation in a twofold manner:

(1) First, it performs density-based clustering. Clusters of data are formed over the dense regions of a point configuration in such a way that the intra-cluster similarity is maximized and the intercluster similarity minimized.

(2) Second, by consecutively assigning values to the bandwidth of the kernel function, the method accomplishes an agglomerative hierarchical clustering. Hence, it starts with the assumption that all the points of the configuration correspond to distinct clusters and successively merges them into a unique one.

The kernel function used for this purpose is the triangular one, \( f_{\text{similarity}} \), defined in Eq. (1):

\[
f_{\text{similarity}}(x, y) = 1 - (1 - \sigma) d(x, y), \quad 0 \leq d(x, y) \leq \sigma
\]

and

\[
f_{\text{similarity}}(x, y) = 0, \quad \sigma \leq d(x, y)
\]

\( f_{\text{similarity}} \) varies with the degree of similarity between \( x \) and \( y \). It decreases until the threshold indicated by \( \sigma \) is reached. The parameter \( \sigma \) corresponds to the kernel bandwidth. The distance between two points \( x \) and \( y \), \( d(x, y) \), takes values from 0 to 1, where 0 denotes dissimilar data while 1 denotes identical data. The sum of the kernel functions constitutes the overall density function, whose graphical representation gives the surface of the landscape. Clusters of similar data can be identified by the local maxima of the density function, but can also be roughly detected visually.

The parameter \( \sigma \) affects the smoothness of the density surface and, by extension, the number of depicted clusters. By setting consecutive values of \( \sigma \), hierarchical clustering is performed, resulting in a nested hierarchy of clusters. More specifically, values of \( \sigma \) smaller than half of the shortest distance between two points lead to the visualization of the lowest level of the hierarchy, where each point corresponds to exactly one cluster. By successively increasing \( \sigma \), the clusters merge into super-clusters, thereby creating the next hierarchical levels. The highest level is visualized when all the points are clustered within the same cluster. The hierarchical levels of the clustering results determine the levels of granularity of the spatialization.

2.2 Labeling strategy

In Geo-Scape, the labels are positioned to the most prominent local maxima and are assigned names according to a dynamic labeling strategy. It is up to the user to select one or more representative variables between the several that describe the observations. Then, each label is assigned a composite name that gives a summary of the variables’ values, as an indicative description of the data nested into the cluster.

The types of variables supported by Geo-Scape are:

(1) The quantitative type of variable contributes its average value to the label name, concerning the observations contained into the corresponding cluster.

(2) The ordinal type of variable contributes its most frequent value to the label name, concerning the ob-
servations contained into the corresponding cluster.

(3) The textual type of variable contributes the value of the observation situated closest to the local maxima described by the label.

For example, suppose that input observations, concerning some geographic sites, are made on the variables Area (quantitative), Landcover (ordinal), and Name (textual). In case the user selects to display information about all the three variables, the label of each cluster will be built from the average value of the variable Area, the most frequent value of the variable Landcover, and the value situated the closest to the local maxima denoting the cluster. For example, the label 1000_Forest_Dadia indicates the location of a cluster that regroups geographic sites with an average area value equal to 1000, mostly covered by forest, and relatively similar to the data with the name Dadia.

Nevertheless, names are reassigned to labels each time the visualized granularity level is changed, because of the merging of sub-clusters into super-clusters or the splitting of super-clusters into sub-clusters. Furthermore, for reasons of clarity, the size of the text fonts increases depending on how much the cluster is populated. Finally, it has to be mentioned that, at the lowest level of granularity, since no cluster has been created yet, the labels directly describe the observations.

2.3 User interaction

Geo-Scape provides an interactive environment within which, users are able to: (1) load input multi-dimensional data, (2) browse the information landscape at a specific level of granularity, (3) change the level of granularity of the representation, namely the smoothness of the landscape, and (4) access information details on demand. Interaction is performed through the prototype’s coordinate views of information: Geo-Scape View, Dendrogram View, and Details View. Coordinate views consist of multiple views linked with each other to show alternative representations of the same information. These views are described as follows:

(1) The Geo-Scape View displays a three-dimensional information landscape where the browsing is made at a specific granularity level and is supported by the operations zooming, panning, and rotating. Zooming consists of changing the spatial resolution dynamically but without changing the granularity level. Panning shifts the view of the landscape up and down as well as left and right, while rotating rotates the view to any orientation. In this way, the user changes viewshed and makes visible parts of the landscape that are potentially occluded.

(2) The Dendrogram View offers an overview of the entire hierarchy of clusters of data. It also highlights the current level of granularity displayed in the Geo-Scape View and thus, lets the user know his position over the whole hierarchy. In order to change the level of granularity, a scroll bar is used that either “generalizes” or “specializes” the landscape. The term generalize describes the operation that augments the landscape smoothness and thus navigates the user to higher levels of granularity. Conversely, the term specialize refers to the exact inverse procedure, which reduces the landscape smoothness and transfers the user to lower levels of granularity by splitting the landscape features into rougher and more complex ones.

(3) The Details View is used to ask for further information in textual form, about clusters’ nested information and about data attribute values. This operation is performed by pointing and clicking on particular clusters (peaks) of the landscape.

In the next section, a demonstration scenario is presented to show how the views provided by Geo-Scape display and handle information.

3 Demonstration scenario

The goal of this demonstration scenario is to show how Geo-Scape helps to discover knowledge into a large set of data, by grouping them into meaningful clusters on the basis of a similarity measure and organizing them hierarchically. Eventually, Geo-Scape could be used to define a categorization schema among a large set of data observations. The scenario uses as input, a set of multidimensional vector data concerning the Greek sites that are protected by the European network NATURA 2000 (Source: European Environment Information and Observation Network, data repository available at http://cdr.eionet.europa.eu/gr/eu/n2000). The data contain 371 observations...
on the following variables: Longitude, Latitude, Minimum Altitude, Mean Altitude and Maximum Altitude. An extract of the data is presented in Table 1 and Table 2. Furthermore, the scenario supposes that the user's task consists of finding groups of data with similar geographical characteristics.

Table 1  Greek sites protected by the European network NATURA 2000

| Site Num | Site description                  |
|----------|-----------------------------------|
| 1        | Dadia forest, Soufli              |
| 2        | Treis Vryses                      |
| 3        | Fengari, Samothraki island        |
| 4        | Mountains of Evros county         |
| 5        | Delta of Evros river              |
| 6        | Delta and west arm of Evros river |
| 7        | Riverside forest of northern Evros river and Arda |
| 8        | Cluster of forests of southern Evros county |
| 9        | Mountainous region of Evros county - Valley of Derios |
| 10       | Mountain Chaidou - Koula and surrounding peaks |

Table 2  Greek sites characteristics

| Site Num | Lon. (º) | Lat. (º) | Alt.Mean (m) | Alt.Max (m) | Alt.Min (m) |
|----------|----------|----------|--------------|--------------|--------------|
| 1        | 26       | 41       | 188          | 614          | 15           |
| 2        | 26       | 41       | 530          | 1034         | 190          |
| 3        | 25       | 40       | 629          | 1600         | -50          |
| 4        | 26       | 41       | 185          | 614          | 13           |
| 5        | 26       | 40       | 1            | 66           | 0            |
| 6        | 26       | 40       | 2            | 32           | 0            |
| 7        | 26       | 41       | 59           | 332          | 0            |
| 8        | 25       | 40       | 235          | 848          | 21           |
| 9        | 26       | 41       | 401          | 1064         | 69           |
| 10       | 24       | 41       | 1262         | 1820         | 683          |

Once the input data have been loaded, Geo-Scape proceeds with the comparison of the observations. The standardized Euclidean distance metric, defined in Eq.(2), is applied as a similarity measure for each pair of data item:

$$
dist(d_i,d_j) = \sqrt{\sum_{a=1}^{n} (x_{ia} - x_{ja})^2 / s_{ja}^2} \tag{2}$$

where $d_i, d_j (i, j = 1, 2, \cdots, 27)$ are a pair of $n$-dimensional data entries to be compared, $x_{ia}$ and $x_{ja}$, the respective values of the variables corresponding to the $a^{th}$ dimension $s_{ja}^2$, the variance of these values, and $n=3$.

Subsequently, once the similarity assessments are computed, Geo-Scape creates the dendrogram resulting from the hierarchical clustering of data and displays it in the Dendrogram View while, at the same time, it creates the information landscape of the data into the Geo-Scape View, at a default intermediate level of granularity. A snapshot of the results is presented in Fig.2.

To this point, it should be noted that there are two kinds of landscape features to look for in interpreting a Geo-Scape Spatialization View: (1) the mountains formed by the regions of high point density, namely the clusters that regroup the points that correspond to similar data, and (2) the linear patterns (straight or curved) formed by elongated complexes of mountains, that reveal the existence of correlation among the characteristics of some subsets of data. In case the data are not related, the resulting configuration forms an irregular “cloud” of points, and the resulting landscape features are uniformly distributed.

In Fig.3, some of the most prominent information revealed by Geo-Scape at the specific granularity level (signified by the dashed line upon the dendrogram in Fig.2(a)) is highlighted. In particular, the landscape reveals: (1) one large and high complex of mountains, denoted with a red circle as the cluster $A$, (2) two other complexes of mountains, also denoted with red circles as the clusters $B$ and $C$, and (3) three linear patterns along which are located elongated complexes of mountains, denoted by red axes upon the landscape.
Cluster $A$ makes the user presume that there exists a densely populated group of sites among the data, with similar characteristics, while the clusters $B$ and $C$, denote the existence of other two, but less populated groups of sites. As was explained above, the axes, or linear patterns, indicate that the corresponding data are so related, that the points have formed a systematic shape, namely the complexes of mountains of elongated form.

Supposing that the user selects to zoom into the cluster $A$ and display the labels containing the variables Minimum Altitude, Mean Altitude and Maximum Altitude of the observations, the Geo-Scape View display area is changed as it is pictured in Fig.4. By noticing the labels, the user can draw the conclusion that the majority of the Greek sites protected by the project NATURA 2000 are areas near the sea because they have an average value for the variable Minimum Altitude equal to 0. In case the user points and clicks on the peak 0_213_523 to ask for a more detailed description of the cluster, the Details View pictured in Fig.5 is triggered and therefore, information details about the relative sites are exposed. Similarly, the user can select to display information about the clusters $B$ and $C$, but also about any other cluster (peak).

Furthermore, the user can select to change the granularity level into a lower one, in order to split the clusters and look at a more detailed view of the data. As the specialized operation is successively used, the granularity level decreases until the landscape finally shows every single data item separately. This way, the user can find sub-categories of data. For example, by displaying labels that describe information about the geographic coordinates of the sites, the user finds which of sites situated near the sea, are situated to neighboring locations.

Conversely, higher levels of granularity merge the clusters and a more general view of the data is then presented. Fig.6 shows how the landscape relief is changing as the granularity of the representation increases.

4 Discussion

Spatial metaphors help the human cognition process\cite{26} and offer spatial affordances, which allow the exploration of information in an intuitive way. However, because metaphors may also imply a “cognitive overhead”,\cite{4} it is preferable that the spatialization method presented is used in case of relatively large and complex data.

However, because of the known limitations of the multidimensional scaling technique, as far as its efficiency in projecting very high-dimensional data...
into low-dimensional dimensional representation spaces is concerned,[27] it is suggested that Geo-Scape is used for the data sets that have a limited number of observation and variables, for example, the input set used in the demonstration scenario.

Geo-Scape tool takes advantage of the third dimension to convey and organize information at different levels of granularity. Nevertheless, opinions vary on the effectiveness of three-dimensional versus two-dimensional hierarchical visualizations. In particular, depth perception and occlusion are two problematic issues often mentioned.[28] The proposed method overcomes the former by making use of linked views. In this way, the users know their depth position over the entire dendrogram while browsing information in the spatialization view. Moreover, linked views stimulate visual thinking by depicting different perspectives of the data.[29] Occlusion is tackled by providing interactive operations like panning and rotating, which can be used to change users’ point of observation and make occluded parts of the landscape visible.

Additionally, the use of color enhances spatialization results and facilitates the innate human ability to recognize spatial patterns. As a result, mountains and hills are effortlessly perceived within the landscape. However, in the near future, we plan to conduct an experiment to test empirically these theoretical foundations and compare our approach to other spatialization methods that manage to depict the granularity of information.

5 Conclusion

This paper discusses the issue of granularity in the context of spatialization approaches. In addition, the prototyping tool Geo-Scape is introduced for visualizing and exploring multidimensional data at different levels of granularity, in the framework of an interactive spatialization environment. More specifically, Geo-Scape produces information landscapes over two-dimensional configurations of points in order to reveal clusters of similar data by means of peaks and organize these clusters hierarchically into different granularity levels through the variation of landscape smoothness.

There are many ways in which we would like to improve our spatialization environment prototype. For example, we are planning to develop a querying mechanism to support the direct retrieval of information. We also envisage the implementation of an extra type of view to isolate part of the information for more focused manipulation. Furthermore, future plans include the elaboration of an ontology integration application based on the proposed spatialization method. This application will aim to determine to what extent different ontologies can be integrated by revealing clusters of semantically similar concepts while taking into account the hierarchical organization of concepts.

References

[1] Skupin A (2000) From metaphor to method: cartographic perspectives on information visualization [C]. Proceedings of the IEEE Symposium on Information Visualization, Salt Lake City, Utah
[2] Demsar U (2006) Data mining of geospatial data: combining visual and automatic methods [D]. Sweden: KTH - Royal Institute of Technology
[3] Card S K, Mackinlay J D, Shneiderman B (1999) Trees [M]. Readings in Information Visualization. Card S K, Mackinlay J D, Shneiderman B, (Eds.). San Francisco: Morgan Kaufmann Publishers
[4] Kuhn W, Blumenthal B (1996) Spatialization: spatial meta-
phors for user interfaces [R]. Reprinted Tutorial Notes from the ACM Conference on Human Factors in Computing Systems in Vancouver, Geoinfo 8, Department of Geoinformation, Technical University of Vienna

[5] Dieberger A, Frank A U (1998) A city metaphor for supporting navigation in complex information spaces [J]. Journal of Visual Languages and Computing, 9: 597-622

[6] Derthick M, Christel M, Hauptmann A, et al. (2003) A cityscape visualization of video perspectives [C]. Proceedings of the National Academy of Sciences, Irvine, CA

[7] Wettel R, Lanza M (2008) CodeCity: 3D visualization of large-scale software [C]. Companion of the 30th International Conference on Software Engineering, Leipzig, Germany

[8] Fabrikant S I, Buttenfield B P (2001) Formalizing semantic spaces for information access [J]. Annals of the Association of American Geographers, 91(2): 263-280

[9] Benking H, Judge A J N (1994) Design considerations for spatial metaphors: reflections on the evolution of viewpoint transportation systems [OL]. http://www.uia.org/ uiadocs/spatialm.htm

[10] Wise J A (1999) The ecological approach to text visualization [J]. Journal of the American Society for Information Science, 50(13): 1224-1233

[11] Boyack K W, Wylie B N, Davidson G S (2002) Domain visualization using VxInsight for science and technology management [J]. Journal of the American Society for Information Science and Technology, 53(9): 764-774

[12] Zavesky E, Chang S F, Yang C C (2008) Visual islands: intuitive browsing of visual search results [C]. Proceedings of the 2008 International Conference on Content-based Image and Video Retrieval, Niagara Falls

[13] Martinez A A, Dolado Cosin J J, Presedo Garcia C (2008) A landscape metaphor for visualization of software projects [C]. Proceedings of the 4th ACM Symposium on Software Visualization, Ammersee, Germany

[14] Skupin A, Buttenfield B P (1997) Spatial metaphors for display of information spaces [C]. Proceedings of the International Research Symposium on Computer-based Cartography AUTO-CARTO 13, Seattle, WA

[15] Kontaxaki S, Tomai E, Kokla M, et al. (2010) Visualizing multidimensional data through granularity-dependent spatialization [C]. Proceedings of the SPIE Conference on Visualization and Data Analysis 2010, San Jose, California

[16] Kohonen T (1995) Self-organizing maps [M]. Berlin: Springer-Verlag

[17] Joliffe I T (2002) Principal component analysis [M]. New York: Springer-Verlag

[18] Mardia K V, Kent J T, Bibby J M (1980) Multivariate analysis (probability and mathematical statistics) [M]. London: Academic Press

[19] Ultsch A (1993) Self-organizing neural networks for visualization and classification [M]. Information and Classification-Concepts, Methods, and Applications. Opitz O, Lausen B, Klar R (Eds.). Berlin: Springer-Verlag

[20] Vesanto J, Himberg J, Alhoniemi E, et al. (2000) SOM toolbox for Matlab 5 [R]. Technical Report A57, Helsinki University of Technology, Finland

[21] Kohonen T, Kaski S, Lagus K, et al. (2000) Self organization of a massive document collection [J]. IEEE Transactions on Neural Networks, 11(3): 574-585

[22] Skupin A (2001) Cartographic considerations for map-like interfaces to digital libraries [OL]. http://www.indiana.edu/visual01/skupin.pdf. Last date accessed 10.2009

[23] Skupin A (2002) A Cartographic approach to visualizing conference abstracts [J]. IEEE Computer Graphics and Applications, 22(1): 50-58

[24] Berkhin P (2006) A Survey of clustering data mining techniques [M]. Grouping Multidimensional Data-Recent Advances in Clustering. Kogan J, Teboulle M, Nicholas C, (Eds.). Berlin, Heidelberg, New York: Springer

[25] Gorg C, Pohl M, Qeli E, et al. (2006) Visual representations [M]. Human-Centered Visualization Environments. Kerren A, Ebert A, Meyer J, (Eds.). Berlin, Heidelberg: Springer

[26] Lakoff G, Johnson M (1980) Metaphors we live by [M]. Chicago: The University of Chicago Press

[27] Skupin A, Fabrikant S I (2003) Spatialization methods: a cartographic research agenda for non-geographic information visualization [J]. Cartography and Geographic Information Science, 30(2): 95-119

[28] Kulyk O, Kosara R, Urquiza J, et al. (2006) Human-centered aspects [B]. Human-Centered Visualization Environments. Kerren A, Ebert A, Meyer J, (Eds.). Berlin, Heidelberg: Springer

[29] Nollenburg M (2006) Geographic Visualization [M]. Human-Centered Visualization Environments. Kerren A, Ebert A, and Meyer J, (Eds.). Berlin, Heidelberg: Springer