Antonio Laganà et al. (Eds.)

LNCS 3044

Computational Science and Its Applications – ICCSA 2004

International Conference
Assisi, Italy, May 2004
Proceedings, Part II

2 Part II

Springer

Antonio Laganà   Marina L. Gavrilova
Vipin Kumar   Youngsong Mun
C.J. Kenneth Tan   Osvaldo Gervasi (Eds.)

Computational Science and Its Applications – ICCSA 2004

International Conference
Assisi, Italy, May 14-17, 2004
Proceedings, Part II

Springer

Volume Editors

Antonio Laganà
University of Perugia, Department of Chemistry
Via Elce di Sotto, 8, 06123 Perugia, Italy
E-mail: lag@unipg.it

Marina L. Gavrilova
University of Calgary, Department of Computer Science
2500 University Dr. N.W., Calgary, AB, T2N 1N4, Canada
E-mail: marina@cpsc.ucalgary.ca

Vipin Kumar
University of Minnesota, Department of Computer Science and Engineering
4-192 EE/CSci Building, 200 Union Street SE, Minneapolis, MN 55455, USA
E-mail: kumar@cs.umn.edu

Youngsong Mun
SoongSil University, School of Computing, Computer Communication Laboratory
1-1 Sang-do 5 Dong, Dong-jak Ku, Seoul 156-743, Korea
E-mail: mun@computing.soongsil.ac.kr

C.J. Kenneth Tan
Queen's University Belfast, Heuchera Technologies Ltd.
Lanyon North, University Road, Belfast, Northern Ireland, BT7 1NN, UK
E-mail: cjtan@optimanumerics.com

Osvaldo Gervasi
University of Perugia, Department of Mathematics and Computer Science
Via Vanvitelli, 1, 06123 Perugia, Italy
E-mail: ogervasi@computer.org

Library of Congress Control Number: 2004105531

CR Subject Classification (1998): D, F, G, H, I, J, C.2-3

ISSN 0302-9743
ISBN 3-540-22056-9 Springer-Verlag Berlin Heidelberg New York

Springer-Verlag is a part of Springer Science+Business Media

springeronline.com

Printed in Germany

Typesetting: Camera-ready by author, data conversion by PTP-Berlin, Protago-TeX-Production GmbH
Printed on acid-free paper SPIN: 11010098 06/3142 5 4 3 2 1 0

Target Data Projection in Multivariate Visualization – An Application to Mine Planning

Leonardo Soto[1,2], Ricardo Sánchez[2], and Jorge Amaya[3]

[1] Department of Systems Engineering, University of Talca
Casilla 747 - Talca - CHILE
lesoto@utalca.cl

[2] Department of Electrical Engineering, University of Concepción
Casilla 160-C - Concepción 3 - CHILE
leonardosoto@ieee.org, risanchez@telsur.cl

[3] Centre for Mathematical Modeling, University of Chile
Casilla 170 - Santiago 3 - CHILE
jamaya@dim.uchile.cl

Abstract. Visualization is a key issue for multivariate data analysis. Multivariate visualization is an active research topic and many efforts have been made in order to find suitable and meaningful visual representations. In this paper we present a technique for data projection in multivariate datasets, named Target Data Projection (TDP). Through this technique a vector is created for each multivariate data item considering a subset of the available variables. A new scalar variable is generated projecting those vectors over a target vector that defines the *direction of interest* for visual analysis. End-users set up target vectors in order to explore particular relationships by means of application meaningful projections. Hence, it is possible to map a combination of multivariate data into one scalar variable for graphical representation and interaction. This technique has proved to be very flexible and useful in mine planning providing valuable information for decision making.

Keywords: Multivariate visualization, mine planning, multiple visualization workspaces, target data projection, human-computer interaction.

1 Introduction

Data representation and manipulation are key activities in problem solving and knowledge discovery. For this reason many efforts have been made to find suitable and meaningful representations of multivariate data [1]. Visualization supports multivariate data analysis by allowing the generation of interactive visual representations aimed to extract knowledge from data. Interactive graphical depicts encourages the users to explore data to find patterns, relationships, and clusters. However, the graphical display of data is constrained to two or three dimensions but data may have dozens of variables. Furthermore, interacting with such a high number of variables may be frustrating.

A. Laganà et al. (Eds.): ICCSA 2004, LNCS 3044, pp. 603–612, 2004.

In this context, one approach is to project the multivariate data over carefully selected data vectors to reduce the dimensionality of the data representation, and still be able to get some information from it. In the statistics domain, several projection methods for multivariate data analysis have been devised. For instance, principal components analysis, factor analysis, and projection pursuit [2]. In general terms, these methods are used to find optimal transformations considering statistical properties of the dataset. Thus, projected representations retain the most informative features of the data. These projections are useful for clustering purposes, supported by orthogonal components. Usually the first principal components are used to define a two dimensional coordinated system, revealing data clusters. However, resulting principal vectors are likely to have no intrinsic meaning for the application domain. Therefore, the expressiveness of data representations is limited when those vectors are used to create a spatial substrate for visualization. It is worth stressing that spatial encodings are perceptually dominant and usually visualization design take advantage of its metric structure [3,4].

Several methods have been developed for multivariate data visualization, including scatterplot matrices, worlds within worlds, parallel coordinates, icon-based display, hierarchical plotting, and pixel based techniques [5]. However, non trivial projections of multivariate datasets have played just a minor role in visualization. Trivial projections refer to those achieving dimensionality reductions by just discarding some data variables. Multivariate data projections methods have been used in text database applications, where documents are treated as points in a multidimensional conceptual space [4,6,7]. Documents are mapped to multidimensional points through text analysis. The associated points are used to classify documents according to its proximity to keyword queries. Tools for text visualization and visual information retrieval has been developed [6,7]. The first attempt to expand the use of data projection was called *data signature* [8]. Data signatures are aimed to capture the essence of large scientific datasets in a compact data vector and to unify different data types. It seems clear that increasing the use of data projection methods in multivariate visualization would result in enhanced analytical capabilities for visualization tools.

In this paper we propose a new projection mechanism for the analysis of quantitative variables in multivariate datasets. Our technique, named Target Data Projection (TDP), relies not on dataset statistics but on application meaningful data vectors for the projection process. Those vectors are referred to as target data vectors because they define a *direction of interest* for visual analysis in the multidimensional space of the dataset. The user specifies a target data vector either by selecting a determined data item directly from the dataset or by defining a vector in terms of interesting data attributes for the analysis (e.g., a particular combination of values for a set of variables). Then, the entire dataset is projected over the established target vector, resulting in a new scalar variable. Generated scalars convey quantitative information about the direction each data item points to, with respect to the target data vector. This information may be extracted with the use of traditional methods for scalar visualization [9]. Besides, the use of classical interaction techniques [3] like brushing or attribute walk is encouraged. Threshold based filtering over this derived variable allows the users to interact in terms of data direction for exploring attribute relations over the dataset, furthering the classical use of range filtering for each variable separately. Similarly, attribute walk is improved with the use of TDP because it is possible to recognize not only data items with similar attribute values, but with similar attribute relations.

Mine planning is a challenging application domain for visualization. Large multivariate datasets describe the mineral resource and establish the starting point for plan design. Strategic, economical, and technical constraints are used to define an optimization model to find an optimal extraction plan [10]. However, the optimization model is far from complete, and many decisions have to be made considering the opinions of several experts. For this reason, an iterative process involving mine plan optimization and discussions of experts from different disciplines takes place throughout plan development. The results obtained in each step should be evaluated by the experts and new constraints would appear. This process relies in the experts to understand complex relations and constraints over a three-dimensional geometry and multivariate data. Hence, visualization tools are of major importance. Our work on TDP was motivated while we were developing a visualization system to support mine planning activities. The focus was in the evaluation of designed plans and planning heuristics considering the available multivariate data.

The rest of this paper is organized as follows: In section 2 target data projection is formally defined. Section 3 describes visualization challenges in the mine planning application domain. Section 4 refers to design and implementation issues in the development of a visualization system for mine planning. Section 5 is devoted to discuss results obtained with TDP, implemented in a visualization prototype. Conclusions and future work are presented in Section 6.

2 Target Data Projection

Reference models provide the foundations to extend visualization techniques and to develop more effective applications [5]. Although work is still required to state an intuitive reference model and notation, *Data State Reference Model* [11] and Stuart Card's model [3] have been of major importance. In the context of these models, *target data projection* would be described as a data transformation. The analytical abstraction behind TDP is a set of multidimensional data vectors defined from the original data table. A derived scalar variable is generated by projecting the whole data table (i.e., its data vectors) in the direction of a particular *target data vector*. Target data vector is intended to be an application-meaningful vector that allows the user to define a *direction of interest* for visual analysis. Finally, original data and generated scalars are assembled resulting in a transformed data table. Exploiting projected variables in the visualization process would result in an improved visual representation or interaction.

2.1 Definition

Let $C_{n \times m}$ be the corresponding matrix for a given data table $\mathcal{T}$ with n data items and m variables, such that c_{ij} element refers to the value of the j-th variable in the i-th data item, $i = 1, \ldots, n$, $j = 1, \ldots, m$.

Now let matrix $D_{n \times m}$ be defined in terms of elements $d_{ij} \in [-1.0, 1.0]$ determined by:

$$d_{ij} = \begin{cases} -1 + 2\frac{c_{ij} - \alpha_j}{\beta_j - \alpha_j} & \text{if } \alpha_j \neq \beta_j \\ 0 & \text{otherwise} . \end{cases} \tag{1}$$

with,

$$\alpha_j = \min_{i=1,\ldots,n} \{c_{ij}\} . \tag{2}$$

$$\beta_j = \max_{i=1,\ldots,n} \{c_{ij}\} . \tag{3}$$

for $i = 1, \ldots, n,\ j = 1, \ldots, m$. Consider the i-th row of D (i.e. the i-th data item) as a m-dimensional vector D_i such that $D_i = [d_{i1}\ d_{i2}\ \ldots\ d_{im}]^t$, $i = 1, \ldots, n$.

Assume a m-dimensional *target data vector* g with coordinates $g_j \in [\alpha_j, \beta_j]$, $j = 1, \ldots, m$. A transformed target vector u is defined with elements u_j given by the expression:

$$u_j = \begin{cases} -1 + 2\frac{g_j - \alpha_j}{\beta_j - \alpha_j} & \text{if } \alpha_j \neq \beta_j \\ 0 & \text{otherwise} \end{cases} \qquad j = 1, \ldots, m . \tag{4}$$

Note that u components are such that $u_j \in [-1.0, 1.0]$, $j = 1, \ldots, m$. Furthermore, a constraint must be imposed in the selection of g to prevent a null target vector u.

$$\|u\| \neq 0 \Longleftrightarrow \exists\, j \in [1, m] : u_j \neq 0 . \tag{5}$$

$$u_j \neq 0 \Longleftrightarrow \left\{ \left(g_j \neq \frac{\alpha_j + \beta_j}{2} \right) \wedge (\alpha_j \neq \beta_j) \right\} . \tag{6}$$

then,

$$\|u\| \neq 0 \Longleftrightarrow \exists\, j \in [1, m] : \left\{ \left(g_j \neq \frac{\alpha_j + \beta_j}{2} \right) \wedge (\alpha_j \neq \beta_j) \right\} . \tag{7}$$

Let a_i be the m-dimensional vector obtained from D_i by projecting to zero every component d_{ij} orthogonal to the target vector u. Thus, a_{ij} coordinates of a_i are determined by:

$$a_{ij} = \begin{cases} d_{ij} & \text{if } u_j \neq 0 \\ 0 & \text{otherwise} \end{cases} \qquad i = 1, \ldots, n,\ j = 1, \ldots, m . \tag{8}$$

then, the *target data projection* of the data table $\mathcal{T}$ in the *direction of interest* determined by the *target data vector* g is defined as the n-dimensional vector s, with elements s_i such that:

$$s_i = \left(Proj_{\hat{u}}\left(\hat{a}_i\right)\right) \cdot \hat{u} = \left(\frac{\langle \hat{a}_i, \hat{u} \rangle}{\|\hat{u}\|^2} \hat{u} \right) \cdot \hat{u} = \hat{a}_i \cdot \hat{u} = \cos\theta . \tag{9}$$

$$\hat{a}_i = \begin{cases} \frac{a_i}{\|a_i\|} & \text{if } \|a_i\| \neq 0 \\ 0 & \text{otherwise} \end{cases} \qquad i = 1, \ldots, n \; . \tag{10}$$

where $\hat{u} = u/\|u\|$,and θ is the angle between the data item vector a_i and the vector u pointing in the *direction of interest*.

The resulting s vector is used for constructing the transformed data table $\tilde{\mathcal{T}}$, associated with the augmented matrix $\tilde{C}$

$$\tilde{C} = \left[\; C \; | \; s \; \right] \; . \tag{11}$$

A slight variant of the method, named *target data projection for unnormalized data vectors* (TDPu) is defined replacing the normalized data vector $\hat{a}_i$ by its unnormalized version a_i. Rewriting equation (6) yields:

$$s_i = \left(Proj_{\hat{u}}\left(a_i\right)\right) \cdot \hat{u} = \left(\frac{\langle a_i, \hat{u}\rangle}{\|\hat{u}\|^2}\hat{u}\right) \cdot \hat{u} = a_i \cdot \hat{u} \qquad i = 1, \ldots, n \; . \tag{12}$$

2.2 Constructing Target Data Vectors

As stated in the previous description, target data vectors are intended to be constructed by the user. The first activity the user should perform is to define a k-dimensional space for data to be projected, $1 \leq k \leq m$. Hence, the original set of m variables is separated in two disjoint subsets. The first subset is named *variables of interest* and is specified by selecting the k important variables for the analysis at hand. For each variable of interest a component should be specified in the target data vector, defining its direction in a k-dimensional space. The second subset is denominated *discarded variables* and is composed of $m - k$ variables excluded from the analysis.

The next step is to compose a target data vector with a particular combination of values in each variable. Components of the target data vector may be specified between two different data ranges: $[\alpha_j, \beta_j]$ or $[-1.0, 1.0]$. If the original data range is selected, then values $g_j \in [\alpha_j, \beta_j]$ are mapped to normalized range values u_j through equation (4).This is useful for defining target vectors by direct selection from the dataset and for specifying its components with critical values for the application. When normalized range is used, values $u_j \in [-1.0, 1.0]$ may be transformed to its original range (see Table 1).

Table 1. Specifying target data vector components.

Variable j	*Normalized Range*	*Original Range*
Discarded	0	$g_j = \frac{\alpha_j+\beta_j}{2}$
Maximum	1	β_j
Minimum	-1	α_j
g_j	$u_j = -1 + 2\frac{g_j-\alpha_j}{\beta_j-\alpha_j}$	g_j
u_j	u_j	$g_j = \alpha_j + \frac{(u_j+1)(\beta_j-\alpha_j)}{2}$

Furthermore, each component of the target data vector corresponding to a discarded variable is set automatically according to Table 1, in order that no direction is defined for this variable.

3 Visualization Challenges in Mine Planning

Mine planning is devoted to the management of processes involved in exploiting mineral resources. Mine optimization and strategic business planning converge to elaborate strategic mine plans for mining projects. Its objective is to develop a strategic mine plan for a given mining project, such that the *net present value* (NPV) of the project is maximized while risk is controlled. From a visualization standpoint, there are several interesting features of the planning process:

1) Handling, exploring and analyzing large multivariate datasets are essential tasks in order to extract valuable information for decision making in plan design.
2) The design process should integrate geological and grade models, geotechnical and operational considerations, economical factors and strategic objectives.
3) Mine optimization is computational intensive. For this reason, heuristics are used to develop the optimized plan. Adjusting planning heuristics requires for fast assessment.
4) Optimization model is far from complete. Discussions of experts should take place in order to recognize critical constraints and to assess the resulting plans. Hence, an iterative process relating mine optimization and experts review is carried out to develop the mine plan.
5) Multidisciplinary interaction must take place to develop realistic plans, including geologists, geo-statisticians, mining engineers, metallurgists and managers.
6) Although high group interaction is required in order to asses design alternatives, there is a lack of efficient and effective ways to communicate ideas in the planning process. Besides, tools are required to present resource information integrated with mine plans to support decision making and plan assessment.
7) The basis for plan design and project analysis is a mineral resource estimate. Consequently there is an associated uncertainty that should be considered.

4 Design and Implementation Issues

Target Data Projection was implemented in a software prototype named SVPM, employing an object oriented approach. Evolving prototypes were developed under C++ using VTK [9]. Close work with mine planning practitioners and the continuous assessment provided for them were key aspects in developing a useful functional prototype. Design was carried out considering the mine as the central object encapsulating data and its graphical representations. Mine object design was organized in three levels: data table, workspace and visual structure.

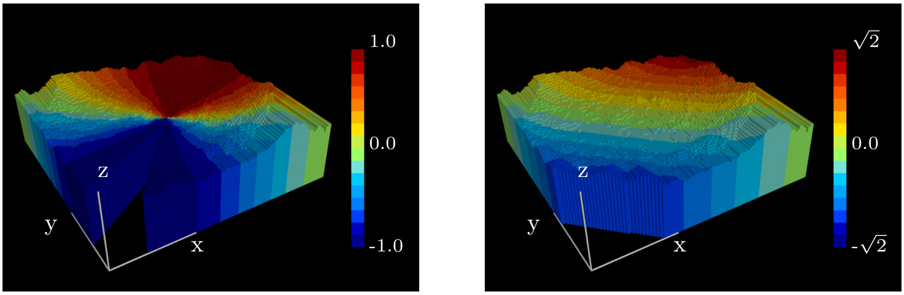
It is worth to mention that for resource modeling and planning purposes, a mine is conceptually divided in many blocks. Hence, a mine has an associated block model, with each block containing multivariate data. At data table level, the block model is used to create a multivariate data table. This level works as a data repository, providing data to construct visual structures in different workspaces. Besides, data table abstraction is considered essential to think about data independently of its intrinsic geometrical meaning and to conceptualize data vectors. At this level, TDP is used to generate user defined projections for expanding visualization alternatives.

Workspace level allows the user to set up a concrete dataset by selecting the variables used to define geometrical coordinates. Hence, spatial substrate is established by choosing one variable for each axis in the visual representation. The remaining variables may be used for color mapping over all the visual structures contained in a workspace. Several workspaces may be used in order to explore different aspects of a dataset. Furthermore, TDP might help, in some measure, to overcome the limitations of a three-dimensional workspace for visualizing multivariate data. This is accomplished by first projecting data vectors comprising several variables, in the direction of a target vector. Then, the projected variable may be used for color mapping or to define one axis in the workspace. Finally, each workspace provides standard view transformations in order to explore its content.

Visual structure level stands for the user to create and manipulate elements in the workspace. Visual structures are generated through visualization techniques, including isosurface extraction, cutting planes, and direct representation of point clouds or mine blocks (blocks representation is supported only for workspaces composed with the original x, y, and z coordinates). User interface elements provide access to each visual structure in a given workspace, allowing the user to set up graphical parameters. Visual structures are arranged according to visualization techniques to create tree-like interface widgets. Standard check boxes are provided to enable and disable the visibility of visual structures in the workspace, either individually or by technique. In addition, threshold filtering is performed in a per visual structure basis. For this reason, a determined visualization technique may be used with the same parameters, but different filter thresholds to create diverse elements in the workspace.

5 Results

SVPM has been used for resource exploration, evaluation of mining plans and validation of planning heuristics with real mine data. High user satisfaction has been achieved through flexible visualization tools and interactive rates for large datasets. In this section, TDP is illustrated for resource exploration tasks. The mine used for illustrations comprises 525.000 blocks and ten scalar variables,



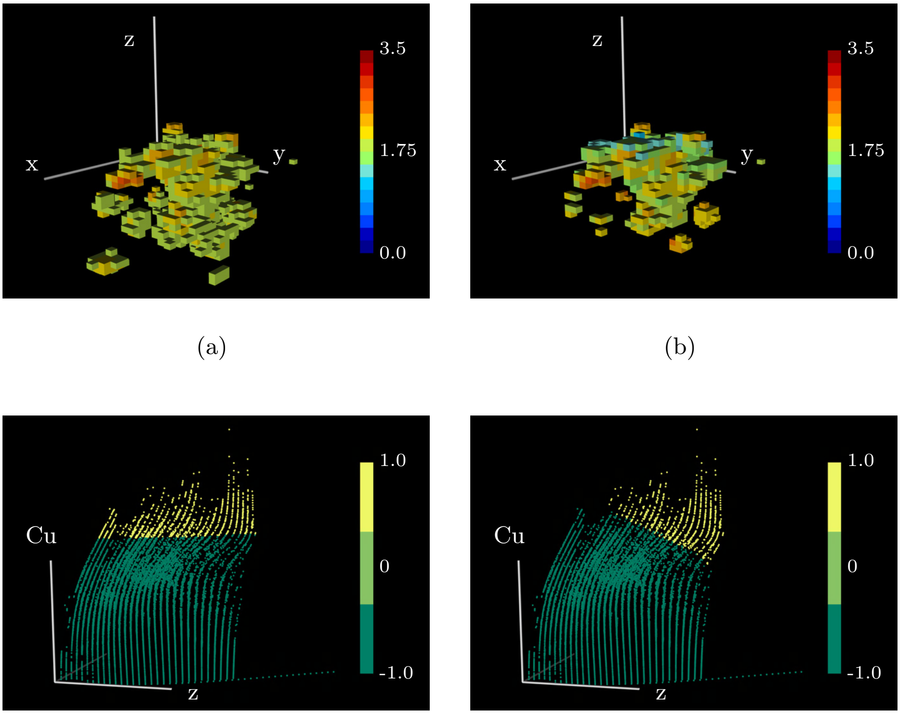
(a) (b)

Fig. 1. Comparing TDP and TDPu in a geographical workspace.

including geographical coordinates, geotechnical parameters, and ore characterization for each block.

Fig. 1 illustrates the differences between TDP (left) and TPDu (right) in a workspace built from geographical coordinates. Classical threshold based filtering is used in order to select only the blocks with projected value in the defined range for the filter. The generated scalars are color mapped over the geometry of the dataset, revealing how to define thresholds for filtering (i.e., how to interact with filters). A target data vector with direction $\{x$-maximum, y-maximum$\}$ was specified by end-users. Fig. 1(a) demonstrate angular filtering with respect to the target vector. This task is achieved using the variable derived through TDP for filtering. In the current image, threshold is set in the range $[\cos 170, \cos 0] = [-0.9848, 1.0]$. It should be noted that, because cosine is an even function, the specified value is considered to both sides of the target vector (this issue is inherent to the projection concept). On the other hand, Fig. 1(b) displays the application of linear filtering in the direction defined by the target vector. This is accomplished using TDPu generated scalars for filtering. In the image, blocks in the range $[-1.0, \sqrt{2}]$ for the generated variable are represented.

TDP may be exploited in more attractive tasks. Ore concentration is a fundamental issue for mining activities. Assume we are interested in finding the spatial distribution of the richest cooper concentrations in the mine. Color coding is set to cooper grade in Fig. 2(a,b). Fig. 2(a) shows blocks above half the range of cooper concentration in the mine. High ore concentrations are revealed, but exploiting mineral resources in the profundity of the mine represents higher costs than extracting resources close to the surface. Note that in open pit mining, extracting a given block requires the preceding exploitation of all the blocks comprising its subsidence cone. This means that to reach a block far from the surface many blocks must be extracted before, and this may take as long as decades. Even rich blocks located in the depth of the mine may present low economic interest, provided that discount rate lowers its net present value (NPV)



(a) (b) (c) (d)

Fig. 2. Using TDP in blocks models and point clouds.

yearly. In this context, TDP may be used to search for blocks in the direction of interesting multivariate combinations. Set a target data vector with direction $\{z$-maximum, cooper-maximum, gold-maximum$\}$ and compute TDP for the dataset. Fig. 2(b) shows blocks in the range $[0, 1]$ for the projected variable P. Compare to Fig. 2(a), note that less blocks are considered interesting in the lower part of the mine. Besides, some blocks with low cooper grade may present high projection values because they are closer to the surface or rich in gold content.

An interesting use of TDP is to split the dataset by using one-dimensional spaces for projections. From expressions (8) and (9), if the target vector u has only one component $u_j \neq 0$ all the projected scalars are going to be either -1, 0 or 1. Consider the previous example, and construct a workspace with coordinates z, cooper and P (projected variable for direction z-maximum, cooper-maximum, gold-maximum). Fig.2 (c,d) shows a point cloud representation corresponding to the block models in Fig. 2(a,b). Fig. 2(c), clearly shows that blocks in Fig. 2(a) were selected according to cooper grade. On the other hand, Fig. 2(d) reveals that threshold filtering in Fig. 2(b) was based on a z-cooper relation.

6 Conclusions and Future Work

Target Data Projection is a flexible, yet easy to implement, tool for visualization systems. It is useful for defining a direction of interest for visual analysis through the specification of a target data vector. The primary application for TDP is to analyze data with respect to vectors composed of maximums and minimums in data ranges of critical variables. This allows application meaningful projections to be generated.

Domains like mine planning, and more recently bioinformatics, demand visualization tools designed with an integrated approach considering both, scientific and information visualization. TDP facilitates this convergence and provides a flexible tool for generating scalars to be visualized with existing techniques. Data table manipulation and workspace definition are key issues in achieving such integration.

Collaborative visualization environments could provide improved support to mine planning activities, from resource analysis to exploitation. SVPM was used in face to face meetings to support discussions, but work should be done to develop a collaborative visualization system for decision making.

References

1. Wong, P.C., Bergeron, R.D.: 30 years of multidimensional multivariate visualization. In: Scientific Visualization - Overviews, Methodologies, and Techniques. IEEE Computer Society Press, Los Alamitos, CA (1997) 3–33
2. Hyvärinen, A.: Survey on independent component analysis. Neural Computing Surveys **2** (1999) 94–128
3. Card, S.K., et al.: Readings in Information Visualization: Using Vision to Think. Morgan Kaufmann Publishers, San Francisco, CA (1999)
4. Ware, C.: Information Visualization: Perception for Design. Morgan Kaufmann Publishers, San Francisco, CA (2000)
5. Ferreira de Oliveira, M.C., Levkowitz, H.: From visual data exploration to visual data minning: A survey. IEEE Transactions on Visualization and Computer Graphics **9** (2003) 378–394
6. Wise, J.A., et al.: Visualizing the non-visual: Spatial analysis and interaction with information from text documents. In: Proceedings of InfoVis'95, IEEE Symposium on Information Visualization. (1995) 51–58
7. Spoerri, A.: Infocrystal: A visual tool for information retrieval. In: Proceedings of IEEE Visualization'93 Conference. (1993) 150–157
8. Wong, P.C., et al.: Data signatures and visualization of scientific data sets. IEEE Computer Graphics and Applications **20** (2000) 12–15
9. Schroeder, W., Mmartin, K., Lorensen, B.: The Visualization Toolkit: An Object Oriented Approach to 3D Graphics. Prentice-Hall, Upper Saddle River,NJ (1998)
10. Turek, C.: Strategic mine planning: Beyond npv - understanding the resource. In: Proceedings of the Whittle North American Strategic Mine Planning Conference. (2000)
11. Chi, E.H., Riedl, J.T.: An operator interaction framework for visualization systems. In: Proceedings of InfoVis'98, IEEE Symposium on Information Visualization. (1998) 63–70