ISBN: 978-972-8939-22-9 © 2010 IADIS

COMBINED REPRESENTATION OF DATA WITH GRAPHS AND HEIGHTMAPS FOR VISUAL ANALYSIS AND SPATIAL DECISION SUPPORT

Vitaly Zabiniako and Pavel Rusakov
Riga Technical University, Institute of Applied Computer Systems
1/3 Meza, Riga, LV-1048, Latvia

ABSTRACT

This paper introduces combined application of network graphs, 2D image processing and spatial information visualization techniques in order to perform preliminary visual analysis in such tasks as planning, optimization and management of spatial networks (e.g., railroad, subway, cargo delivery, etc.). Application of proposed methods is also possible for any kind of analysis that involves exploring spatial relationships between multiple objects that form topological network structure. Definition of general problems in this field is being presented with description of appropriate existing methods for solving them. Theoretical background behind the concept of proposed method is revealed along with evaluation of its application in a case study. Conclusions about current achievements and potential development of this method in the future are also presented.

KEYWORDS

Topology, space, analysis, decision, graph, heightmap.

1. INTRODUCTION

Solving of structured or semi-structured spatial problems (for example – in communication networks) is so hard that even nowadays it still pushes forward development of such complex knowledge domains and tools as logistics, supply chain management and Decision Support Systems (DSS). The latter concept is in connection with interactive, computer-based systems designed to support a user with guidance and necessary information in making land use decisions taking into consideration multiple aspects such as transportation, water management, demographics, agriculture, climate, employment and many other factors. Each such system requires a set of sophisticated problems solving algorithms, strategies and decision models in its inventory. A good example would be planning of extension of current network infrastructure with adding new elements (stations, hubs, warehouses etc.). Of course, in a real application according analysis will take into consideration dozens of factors such as demand trends, existence of alternative solutions, etc. Usually this will result in solving of multicriteria optimization task.

Still, before these refined methods will be put into the action, there is always a need for preliminary analysis that is performed by a human in order to identify general characteristic and key aspects of the spatial problem under investigation. At this preliminary stage highly detailed characteristic of the network are pretty much irrelevant because the main emphasis is put on such primary properties as the topology of the network, geographical (taken to the higher abstraction level – geometrical) position of its elements and individual associated quantitative values (throughput \ income \ number of passengers \ transactions, etc.).

Let's consider the following general task: knowing the spatial locations of discrete network nodes, relationships among them and general useful property of each node and\or relationship, find according quantitative property for the new network element with desired location. This scenario is applicable to the situation when there's a need to build new railroad station in a certain region and the user wants to estimate its expected usage rate and thus – income. Intuitively we understand that new station in a region with higher railroad usage activity will bring more profit than the station in a region with lower activity, but how to obtain more formal-based answer?

Common sense is averaging \ interpolation \ extrapolation of values of geometrically close network elements. But there's another problem – definition of "closeness". Even with adjustable sampling radius, we get discrete set of probes without being sure that it will provide enough precision for finding the average. Ideally we want to construct continuous function

$$X_i=F(L_i) \qquad (1)$$

where X_i – target value;

L_i – given location;

F – continuous function for the transformation $L_i \rightarrow X_i$.

Another decision type might be as follows: knowing the same information as in previous example, identify a set of separate regions $R_1=\{L_1, \ldots, L_i\}$, $R_2=\{L_2, \ldots, L_j\}$, … $R_n=\{L_n, \ldots, L_k\}$ with corresponding $X_i > \alpha$. Semantically this is equivalent to clustering obtained target values X_i and filtering regions of interest with threshold α.

The general goal of this research is to propose an approach that might help in solving mentioned interpolation and clustering problems during preliminary spatial analysis. There are four subtasks being defined: 1) to present theoretical base for construction of proposed approach; 2) to demonstrate the usage of this approach in a real case study; 3) to outline main principles of existing solutions; 4) to make conclusions and to provide information about planned enhancements of the proposed method.

2. CONSTRUCTION OF CONTINUOUS FUNCTION BY CONVERTING GRAPH INTO HEIGHTMAP

The concept of finding useful function, which is described above, includes 5 general steps:

1. Visualization of network graph in three-dimensional Cartesian coordinate system using two dimensions for mapping geometric locations of the elements, while using the third dimension for mapping known quantitative values attached to each graph node or edge.
2. Drawing this spatial graph into grayscale, considering "weight" of its parts.
3. Extending grayscale graph image to ordinary heightmap.
4. Performing smoothing of the resulting heightmap with two-dimensional image filters.
5. Combining visualization results of steps 1 and 4 for the purposes of visual analysis.

The general schema of data processing flow is presented in Fig.1.

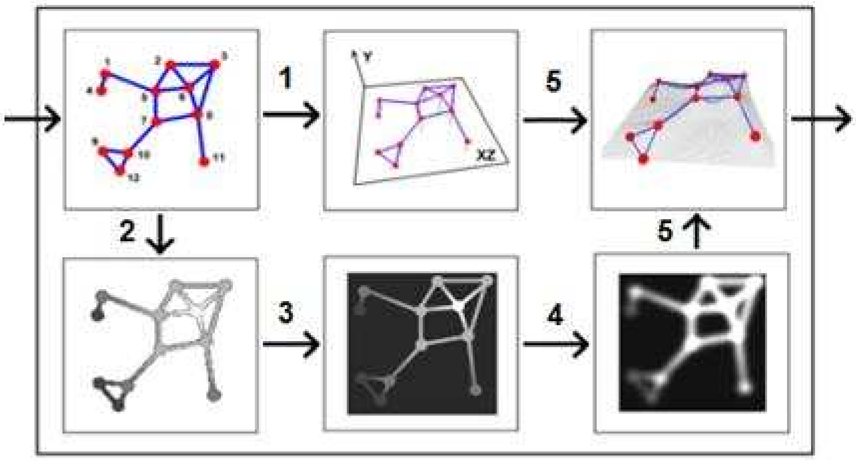

Figure 1. Data processing flow

This process will be described in details below. The data set that will serve for demonstration purposes is presented in Table 1 (quantitative characteristics) and Fig.2 (topological data captured in form of a graph).

Table 1. Quantitative characteristics of demonstration data set

Node number	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Associated value	67	85	82	28	72	99	80	76	26	63	64	22

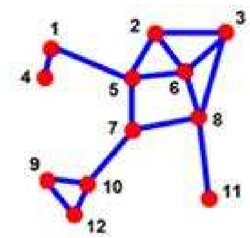

Figure 2. Topology of demonstration data set

At the first step the graph body is prepared for visualization in three-dimensional space so that the mapping of the topology occurs in the *XZ* plane (considering right-hand Cartesian coordinate system) and mapping of the quantitative data is associated with *Y* axis (see Fig.3, part 1). The mapping is based on relative scale so that the maximal value of input data (in this case – *99*) is associated with predefined *Y* value (in this case *255* units of the coordinate system that correspond to pure white RGB color value – see below).

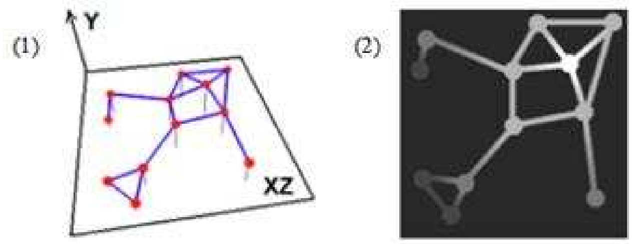

Figure 3. Initial steps of data processing

The second step involves converting initial graph into grayscale image that will form the basis of the heightmap. This time, same quantitative data must be mapped to linear grayscale gradient with minimal and maximal values corresponding to pure black and white colors respectively (RGB component values *0* to *255*). Then the two-dimensional image of the graph must be produced. If associated values are known not only for nodes but also for edges or even individual edge segments, these may be used explicitly. Otherwise, as in given example, values for individual pixels of each edge must be gradually interpolated between nodes. For this purpose each edge might be substituted with gradient bar with same length, width and appropriate rotation. During the third step the complete heightmap is formed by filling remaining unoccupied area with black color (semantically this represents absence of any usable value in empty regions) – see Fig.3, part 2.

Although it is possible to visualize heightmap at this stage immediately, there is little use of it, as it has very coarse distinct borders between regions. It is necessary to smooth these borders and that is when two-dimensional image processing starts to play important role in the fourth step. Applications effects of three common processing techniques are being used in this work, namely – Gaussian low-pass blur, contrast and brightness.

Blurring is known as the technique for smoothing raster images. The mathematical model of changing blur level is based on the following well-defined function that is widely used not only in image processing but also in signal processing and statistics [12]:

$$G(X) = \frac{1}{\pi\delta^2} e^{-\frac{|X|}{\delta^2}} \qquad (2)$$

where X – length of the vector from the origin (in case of *2D* image $X = \sqrt{x^2 + y^2}$);

δ – standard deviation of the Gaussian distribution.

By constructing a surface in two dimensions (with given δ) and performing sampling, it is possible to obtain convolution matrix *B* that defines averaging coefficients. Then each pixel of the image is modified in a loop, by taking into consideration current RGB color values of its neighbor pixels and color of the current pixel according to matrix *B* by the following algorithm [19]:

$$Color' = \sum_{i=1}^{|B|} Color_i * B_i \qquad (3)$$

Stronger blur effect (and thus – better averaging) can be achieved either by increasing number of elements in the matrix *B* or by applying same matrix several times for the entire image.

Contrast for individual pixel is expressed by difference of its brightness and average brightness of the entire image [4]. Thus, adjusting of the image contrast can be achieved as follows:

$$Color' = Color_M + (1 + \frac{\varepsilon}{100}) * (Color - Color_M) \qquad (4)$$

where

$$Color_M = \frac{\sum_{i=1}^{n} Color_i}{n} \qquad (5)$$

$\varepsilon \in [-100;100]$ – contrast correction parameter;

n – number of pixels in the image.

Brightness corresponds to the luminescence amount of individual pixel. Adjustment of individual RGB component of pixels with parameter α is trivial:

$$Color = \begin{cases} 0, & if \quad Color + \alpha < 0 \\ Color + \alpha, & if \quad 0 \le Color + \alpha \le 255 \\ 255, & if \quad Color + \alpha > 255 \end{cases} \qquad (6)$$

The key point in application of these techniques for achieving desired results is the order in which these are applied. Let we have the original grayscale image (Fig.4 – part 1). In case the construction of previously mentioned continuous functions is needed, the first step should be blurring (Fig.4 – part 2). Blurring effect must be relatively strong, as we want to remove coarse borders and extend these further from their original position. By experimenting with different input data and making empiric evaluations we concluded that adjusting δ so that blur radius is equal to about *4% – 5%* of averaged image height and width is enough.

After application of blurring, initial data is diffused with intensity of neighborhood regions, so in the next step it is necessary to increase contrast of the heightmap by applying according technique (Fig.4 – part 3). The recommended ε value in this case is about *70%*. Application of these two operations forms the first data processing iteration at the fourth step of the proposed algorithm. The algorithm may either proceed to step 5 or perform the next smoothing \ contrasting iteration of same kind (Fig.4 – parts 4 and 5).

If it is necessary to perform visual clustering to distinguish areas with high value intensity, another kind of two-step iteration is required. It consists of intensive contrast adjustment (ε is close to *100%*), followed by adjustment of brightness that sets the threshold α that is mentioned in the definition of the clusterization problem in chapter 1 (Fig.4 – part 6).

So, all chain of application of mentioned transformations and achieved result are presented in Fig. 4.

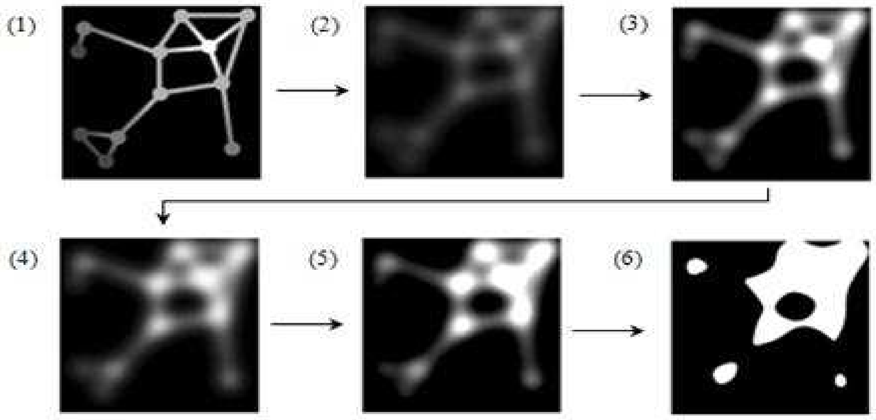

Figure 4. Chain of image processing transformations

In order to evaluate statistical effect of application of these filters, the arithmetic mean and standard deviation of resulting color intensity were calculated for each step – refer to Table 2 and graph in Fig. 5.

Table 2. Color intensity table

Step	**1**	**2**	**3**	**4**	**5**	**6**
Arithmetic mean	22.32	21.07	43	42	48	60
Standard deviation	57	26	68	61	79	107

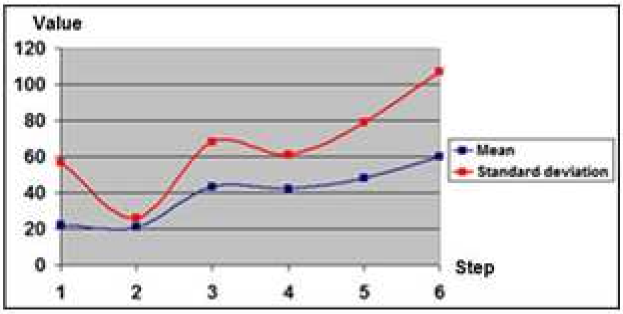

Figure 5. Color intensity graph

During the fifth, final step of the algorithm, the function surface is visualized together with graph body in three dimensions that allows for the user to obtain general combined vision.

Construction of the surface is as simple as placing an array of vertices in space that conforms to the characteristics of each individual heightmap pixel (position and color intensity) and connecting these in any suitable mode – with individual lines \ line strips \ quad polygons, etc. – refer to Fig.6.

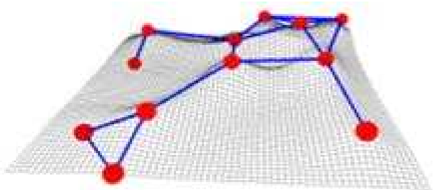

Figure 6. Combined visualization of graph body and function surface

3. A CASE STUDY: ANALYSIS OF RAILROAD INFRASTRUCTURE IN LATVIA

By utilizing such tools as dynamic three-dimensional camera, color management and similar visualization techniques, it is possible to represent distribution of useful values in convenient homogenous form.

The structure of railroad of Latvia was used as real dataset for analysis using proposed method. Official map of interconnected Latvian cities and routes [16] was chosen as the input – radius of nodes depicts averaged amount of passengers per each station (Fig.7 part 1). This map was converted to appropriate heightmap (Fig.7, part 2) and processed by utilizing above-mentioned procedure (Fig.7, part 3). The last heightmap visualization step with the distribution of averaged activity is shown – see Fig.7, part 4 for overall perspective; part 5 for region close up.

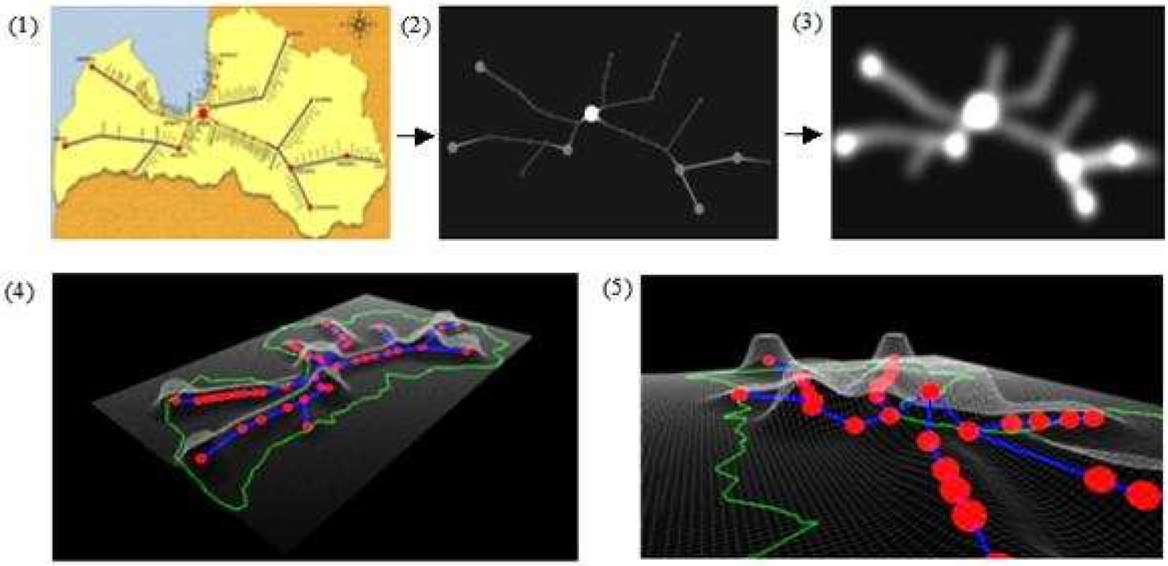

Figure 7. Railroad network of Latvia

In order to locate suitable regions for further expansion, the map of population density in Latvia (based on [3]) was used (see Fig.8, part 1).

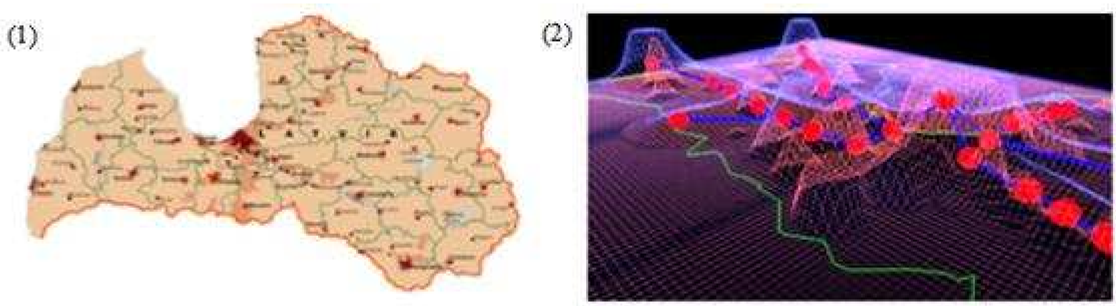

Figure 8. Combination with averaged population density

According heightmap was extracted and added to the previous result using visual separation with surface color (activity – blue, density – red). In this case color image was desaturated so that only pixel luminance defined the input. As a result, it is possible to visually identify regions with high population level and low railroad coverage (see Fig.8, part 2). Although these images provide useful information about infrastructure of Latvian railroad system and its demographic environment, there was no intention perform real optimization, as it depends on many other (including historical) relevant aspects and constraints that are not taken into account. As it was stated in introductory part of this work, the main emphasis of the method is put on preliminary analysis during early planning stage.

4. RELATED WORK AND COMPARISON

The concept of Spatial Decision Support (SDS) emerged in the 80-th and stays as active topic since then. It also relates to the concept of Geographic Information Systems (GIS) as described in [11]. The author of this paper argues that GIS are general in their nature, while SDS are more focused on the rules of specific knowledge domains that utilize spatial characteristics of objects. Considering that in most cases making of spatial decisions is based on multiple criteria simultaneously it is of great importance to summarize according existing methods [14], propose methodological improvements of these [5] and define future trends as in [2].

The problem of construction of continuous functions that may serve for the analysis purposes is closely related to construction of Digital Elevation Models (DEM) that are at the heart of modern GIS and other topographic systems [8]. According paper explores different algorithms for interpolation for a set of discrete points obtained during measurement and offers improvements that aim to increase the accuracy of the result.

Solving of the general (multidimensional) interpolation task is very general problem that has been researched for generations, resulting in application of techniques ranging from basic linear averaging to much more complex nonlinear methods. For example [13] use directional filter banks; in [17] Laplacian Pyramid is used for decomposition and prediction of local high frequency components; [15] proposes an iterative method for smoothing based on level set theory and curves of constant intensity; in [20] the use of bilinear interpolation and correction of the error with the interpolation error theorem are being used, etc.

Clustering is another important concept in the business analysis. Detecting of dense semantic relationships among certain entities may influence ongoing management strategy and even trigger development of new business rules crucial for increasing income. Authors of [6] provide an outline of methods for clustering of spatial networks based on a graph model. Paper [1] introduces another original approach to clustering problem that is able to perform efficiently without assumptions about distribution of the data and input of the thresholds. Authors of [9] propose a method for density-based notion of clusters that is designed to discover clusters of arbitrary shape.

Visualization of data as a continuous surface is described in [7] along with custom smoothing techniques and modifications for produced triangular mesh. The authors of [12] use same Gaussian blur for the first step of data visualization, however according to this approach, data is presented in two-dimensional space using rainbow gradient for visual distinction of regions with different data density.

It is necessary to note that opposite solutions are also known in an attempt to generate topological information in form of graphs from spatial data encoded in heightmaps – refer to [18].

In order to evaluate interpolation results produced by our method in comparison with other approaches, additional experiments were performed. A set of ten test samples has been chosen on demonstration data defined in chapter 2 (see Fig.9, part 1). Target values X_i obtained with our approach were compared with results generated by "Dataplot" software (refer to description of according interpolation method in [10]) running on the same dataset. Comparison results are presented in Fig.9, part 2.

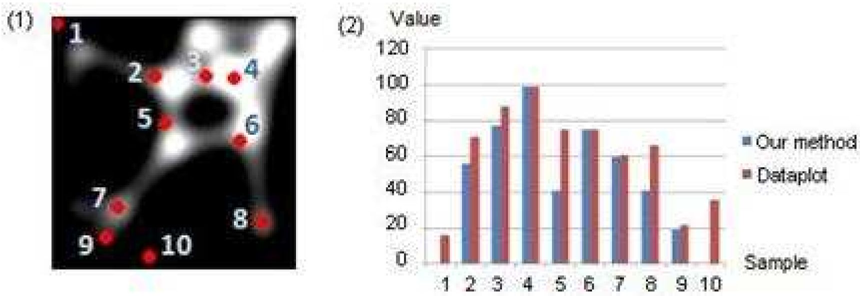

Figure 9. Comparison of methods with a set of samples

As it might be perceived from the chart, correlation exists between both results. The main distinction is that our method "fadeouts" obtained values faster, although this can be flexibly controlled by amount of brightness and contrast applied during image processing.

5. CONCLUSION

The traditional approach in solving general analysis problems, that tend to be topological and numerical simultaneously, requires application of complex mathematical methods – both continuous and discrete. In case the data under investigation can be represented with graph elements and associated quantitative properties, authors of this article propose to make a shift to the methods of computer graphics (based on previous researches in graph visualization and according analysis as in [21], [22]) by combining three-dimensional visual representation of data with its processing with standard filters for two-dimensional images that can be found in most modern raster graphics editors. Even such common techniques as blurring \ contrast \ brightness might serve as effective tools for preliminary analysis and making spatial decisions if applied in combination and in correct order. The main contribution of this work is the proposed approach of combination of topology of graphs and heightmaps that treats previously mentioned problems in a single elegant way by defining several strategies for application of well-known processing techniques for general analysis scenarios (“blur – contrast” for interpolation, “contrast – brightness” for clustering).

According to the experiments, blurring decreases both the mean and standard deviation of the intensity while contrast increases these characteristics. By applying the described chain of smoothing and contrasting iterations it is possible to obtain distribution of values that ensure smoothly averaged surfaces.

It is important to note that the proposed approach is suitable for objects geometrically located in two-dimensional space only – the third dimension is reserved for processing of data, associated with these objects. Still, there is no such strict limitation for number of dimensions of associated data.

During further development of proposed approach it is planned to increase number of attributes. This can be done by representing these with such properties as dynamically calculated color of the heightmap grid, width of its lines, transparency of filling polygons, etc. In this case information storing and processing can be based on color images that use RGB / RGBA channels for individual attributes, although this approach is limited with number of supported data dimensions.

ACKNOWLEDGEMENT

This work has been supported by the European Social Fund within the project "Support for the implementation of doctoral studies at Riga Technical University".

REFERENCES

Anders, K., 2003. A Hierarchical Graph-Clustering Approach to Find Groups of Objects. *In ICA Commission on Map Generalization, 5th Workshop on Progress in Automated Map Generalization*. Paris, France.

Ascough, C. et al, 2002. Multicriteria spatial decision support systems: overview, applications, and future research directions. *Proceedings of Integrated Assessment and Decision Support (iEMSs)*, Lugano, Switzerland, pp. 175-180.

Best Country Reports, 2007. *Population Density Map of Latvia*. World Trade Press, Petaluma, USA. http://www.bestcountryreports.com/Population_Map_Latvia.html

Bezryadin, S. et al, 2006. *Adjusting contrast in computer software* (“Preobrazovanije kontrasta v programmnom obespecenii” (in Russian)). KWE International. Inc., San Francisco, USA. www.kweii.com/site/color_theory/Contrast100_ru.pdf

Chakroun, H. and Benie, G., 2005. Improving Spatial Decision Support Systems in Natural Resources and Land Management. *Applied GIS*, Vol. 1, No. 1.

Demir, et al, 2008. Clustering spatial networks for aggregate query processing: A hypergraph approach. *Information Systems*, Vol. 33 No.1, pp. 1-17.

Dieter, W., 2009. *Visualization of polygon-based data as a continuous surface*. Manuscript. http://www.wdrase.de/VisualPycnoInterEngl.pdf

Droy, G., 2008. Improving the accuracy of Digital Terrain Models. *Babes-Bolyai Studia Universitatis magazine INFORMATICA*, LIII (1), pp. 65-72.

Ester, M. et al, 1996. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*. Portland, USA, pp. 226-231.

Franke, R., 1982. Smooth interpolation of scattered data by local thin plate splines. *Comp. Maths. Appls*. Vol 8, pp. 273-281.

Keenan, P., 2006. Spatial Decision Support Systems: A coming of age. *Control and Cybernetics*, Vol. 35, No. 1, pp. 9-27.

Liere, R. and Leeuw, C., 2003. Graphsplatting: Visualizing graphs as continuous fields. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 9, No. 2, pp. 206–212.

Lu, X. et al, 2003. An efficient directional image interpolation method. IEEE International Conference on Acoustics, Speech, and Signal Processing, Proceedings, Vol. 3, pp. 97-100.

Malczewski, J., 2006. A GIS-based multicriteria decision analysis: A survey of the literature. International Journal of Geographical Information Science, Vol. 20, No. 7, pp. 703-726.

Morse, B. and Schwartzwald, D., 2001. Image magnification using level-set reconstruction. Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Kuai, USA, Vol. 1, pp. 333-340.

"Pasažieru Vilciens", 2009. Railroad map. Latvijas dzelzeļš, Latvia. http://www.pv.lv/?cat=298

Takahashi, Y. and Taguchi, A., 2002. An enlargement method of digital images with the prediction of high-frequency components. IEEE International Conference on Acoustics, Speech, and Signal Processing, Proceedings, Orlando, USA, Vol. 4, pp. 3700 – 3703.

Weinkauf, T. and Gunther, D., 2009. Separatrix Persistence: Extraction of Salient Edges on Surfaces Using Topological Methods. Computer Graphics Forum, Vol. 28, No. 5, pp. 1519-1528.

Wright, R. and Lipchak, B., 2004. OpenGL SuperBible (third edition). Sams, Canada, USA.

Youngjoon, C. and Seongjai, K., 2007. The Error-Amended Sharp Edge (EASE) Scheme for Image Zooming. IEEE Transactions on Image Processing, Vol. 16, No. 6, pp. 1496-1505.

Zabiniako, V. and Rusakov, P., 2007. Analysis of Visualization Problems of Graphs and Models of Graphs. Scientific proceedings of Riga Technical University, Computer Science, Applied Computer Systems. Riga, Latvia, pp. 138-148.

Zabiniako, V. and Rusakov, P., 2009. Definition of General Requirements for Graph Visualization Software. *Scientific proceedings of Riga Technical University, Computer Science, Applied Computer Systems*. Riga, Latvia, pp. 168-179.