RELEVANCE OF VISUAL EXPLORATION FOR STRENGTHENING SPATIAL THINKING & SPATIAL KNOWLEDGE EXPLORATION

Abdulvahit Torun(*), Şebnem Düzgün (**)

(*) General Command of Mapping, Cartography Dept., 06100 Cebeci, Ankara, Turkey
(**) Middle East Tech. Univ., Geod. & GI Tech. Dept, Ankara, Turkey -
atorun@hgk.mil.tr; duzgun@metu.edu.tr

Commission II, WG II

KEY WORDS: Geo-spatial Visualization, Visual Exploration, Spatio-temporal Data, Discovering Spatio-temporal Patterns, Knowledge Discovery, Spatial Data-mining, Spatial Data Analysis, Visual Thinking.

ABSTRACT:

Tremendous amount of spatial data reside in operational or legacy data stores of public and private institutions. These databases contain topographic maps, aerial photos, satellite images, medical data, laser/lidar scanner data, video images among others. In addition to spatially referenced data, there are links from spatial objects to non-spatial data such as census, economic, security, and statistical information. It is costly and often unrealistic for users to examine spatial data in detail and search for meaningful patterns or relationships among huge amount of data. Spatial data mining (SDM) aims to automate such a knowledge discovery process in large databases along with visual exploration techniques.Visual exploration allows user to interact with huge spatial datasets to recognize spatial distributions and relationships and to extract meaningful information by facilitating visualization of spatial patterns and processes, and other cognitive skills such as visual thinking, visual comparison. This paper introduces visual exploration methods complementing spatial data mining methods, visual representations of knowledge exploration on historical ship accident data at Istanbul Strait, crime data in Ankara and earth quake data of Turkey. These methods are used to visualize and explore spatio-temporal data, to support visual thinking and knowledge exploration, to reveal unknown relationships.

1. INTRODUCTION

Web publishing, democratization and localization of information by disseminating via internet increased interest of institutions, companies and individuals towards geo-spatial information mapping, visualization and analysis. In addition to the geo-spatial visualization, content visualization, data mining, and knowledge discovery of non-spatial data are depicted as map metaphors to increase visual thinking and cognition.

Legacy or operational spatial databases or non-spatial (geographic) databases, which are also somehow linked to a spatial database comprising census, economic, statistics, intelligence are subject to analysis for decision and policy making. Spatial databases reside terabytes of spatial data that may be obtained from topographic maps, aerial photos, satellite images, medical equipments, laser/lidar scanners, video cameras among others in public and private organizations.

Data on its own has no value. Without simple visual ways to integrate, display and analyze, it is possible to end up with massive amounts of data but no information. However, it is costly and often unrealistic for users to examine spatial data in detail and search for meaningful patterns or relationships among data. Spatial data mining (SDM) aims to automate such a knowledge discovery process in large databases along with visual exploration techniques for correct communication.

Due to limitation of human cognition and complexity of tasks, filtering un-important information and focusing on relevant information is basic power of spatial visualization. In order to allow the human interactively get insight into the data, the techniques such as thematic mapping, multi-variate visualization, brushing, attribute linking and non-spatial visual exploration methods are employed to recognize patterns, to reveal relationships between data attributes, to identify global structures and finally draw instant conclusions directly from spatial data.

Visual data mining consists of a series of visualization and visual exploration methods, connected in a sequence to present the data in some visual form for strengthening visual thinking and knowledge exploration

In this study, visualization techniques are introduced within context of spatial thinking, visual exploration and visual data mining as an independent part and as a visualization tool of SDM. Then, an application is implemented on historical data of oil transportation and ship accidents at Istanbul Bogazi (Istanbul Strait) to discover spatial patterns among the data and its environment.

The second chapter gives basic concepts on SDM, visual exploration and GIS as a data mining and visualization tool. Chapter three comprises application, experiences and results. The last chapter discusses the final ideas and conclusions.

2. VISUAL EXPLORATION AND KNOWLEDGE DISCOVERY

Determining patterns, behaviours, trends, occupation and successive position of a phenomenon in the field based on their distribution, association among each other and their interaction with the environment are the basic issues of spatial knowledge exploration. Conventional approaches may not be effective and efficient for handling such a complex task which requires integration of heterogeneous data, modelling, analyzing, visualization and human perception and judgment.

Visual exploration as a non-separable part of spatial knowledge discovery techniques are methods to model, visualize and explore spatio-temporal data to support visual thinking to reveal unknown relationships.

Geo-spatial visualization by means of cartographic techniques transforms raw geographic data into visual representations that are both attractive and useful. The geo-spatial sciences, like geography, cartography, and surveying have developed numerous methods and techniques to capture, process and visualize geographic information. However, research and application in information visualization rarely makes reference to geographic or cartographic research. The visualization should not necessarily be like a map but mapping metaphor is seriously used in information visualization based.

Recent progress in scientific and engineering applications has accumulated huge volumes of high-dimensional data, stream data, unstructured and semi-structured data, and spatial and temporal data. This complexity may require multi-level spatio-temporal data mining and visualization. Usually, techniques for exploring spatio-temporal knowledge are put in a range from visualization to data mining which may be classified as follows;

- Spatio-temporal visualization
- Spatio-temporal analysis and exploratory visualization
- Spatio-temporal data mining

2.1 Spatial Thinking and Mental Model of Reality

Human usually construct a mental model of reality, which is an abstract form of simplified versions of real systems. Spatial information enables a better grasp of spatial problems due to better visualization.

With visual exploration techniques, the result is obtained faster and with a higher degree of confidence, because the exploration is intuitive and doesn't require understanding of complex mathematical algorithms. It is effective when little is known about the data and when the exploration tasks are vague. The user only know how to react rather than knowing techniques behind the tool by developing new measures and models. Visual exploration techniques may include uncertainty in data and vagueness of tasks.

Visualizing spatio-temporal data in new ways and presentation of alternative perspectives may stimulate exploration of new hypotheses for spatio-temporal data mining. The greatest contribution of visualization to the process of scientific thinking is liberating the brain from the fundamental activity of information retrieval and manipulation required to produce an image, thereby allowing the brain to devote its time and energy to higher levels of analysis and synthesis (Friedhoff and Benzon 1989, McCormick *et al.* 1987).

2.2 Spatial Data Visualization

Cartography –visualization of geo-spatial data- does not involve only knowledge representation, but also involve knowledge abstraction which is based on a model for visual exploration. Visual knowledge abstraction requires further visualization techniques in compliant with spatial analysis, modelling and knowledge discovery (Zucker et.al. 2000).

Spatial data visualization methods and spatial data analysis techniques are employed to explore spatial data. Techniques, which have no visual view are not so helpful to identify spatial patterns (Fotheringgham et.al. 2000). Boxplots, histograms, parallel coordinate plots, projection pursuit, univariate/bi-variate scatter plots and leaf plots are graphical views of spatial data which are not depicted in map format in 2D or 3D.

Cartographic visualization is the most effective way of communicating information about the location and spatial characteristics of the natural world and of society. Choropleth maps and cartograms are spatial representation in 2D and 3D. Choropleth map is visualization of area –lattice or enumeration- data regarding an attribute in gradual colours. Resizing the area proportional to an attribute is depicted as cartograms. Linked plots are interaction tools to link and visualize different views of spatial data. Slicing helps the user to understand the patter while one or two of spatial/temporal/attribute is fixed. The result may be a static or dynamic (spatial/temporal animation) map. There are various software packages –stand alone or available on the internet- that particularly make use of geo-spatial visualization techniques for non-geographic information such as SPIRE (Spatial Paradigm for Information Retrieval and Exploration) and Viscovery SOMine.

Visualization techniques concerning time variation, 3-D and interactive navigation are essential parts of visual exploration to support perception. Including dimensionality such as spatial, temporal, symbolic/attribute into visualization extends the ability of representation to incorporate capabilities like dynamic visualization and animation in addition to 2D, 3D traditional visualization techniques.

Visualizing large datasets as 3D and dynamic spatial animation is particularly a challenging task because of computational and representational difficulties (Friedhoff et.al 1989). Computational site is handled by computer graphics based on powerful open standards such as openGL. Although, todays GIS software enable efficient and effective 3D and animation tools, there is still difficulty viewing information. Several visualization methods such as windowing, fish-eye views, link inheritance, tree condensing, and link typing are employed in hyper media visualization.

2.3 Considerations in Visual Exploration and Information Processing

Exploratory visualization comprises visualization and visual data mining techniques. Cartographic visualization allows user to interact with huge spatial datasets to recognize spatial distributions and, relationships and extract meaningful information by facilitating comparison, conceptualizing spatial patterns and processes, and other cognitive skills by making use of the visualization enabled systems.

The difficulty of visualizing high information spaces on two dimensional displays severely prevents perception of richness of data and complexity increases as data and its dimension grows. There are several issues to be considered in spatial visualization and visual exploration. Some of them are dimension reduction, sampling size, scale, similarity measures, data transformation, dimension reduction, distance and generalization.

Dimension Reduction:

When data has more than 3D, visualization prohibits visual mining. There cartographic techniques such as generalization, aggregation, summarization procedures should be used to simplify data.

Sampling Unit and Scale (Level of Detail):

Sampling of geographic space and granularity of information space subject to spatial analysis exhibits certain scale-dependent structural characteristics. Thus the scale dependency of geographic data appears relevant both analysis and visualization point of view to preserve geographic reality. Level of detail (in broad sense resolution and granularity and scale) is one of the main issues to be determined for both analysis and visualization in ESDA. Moreover, both spatial and radiometric resolution stand out main issues of digital data processing in remotely sensed raster imagery.

Similarity:

Selecting the similarity model is a basic problem for spatial data analysis such as auto-correlation, clustering, classifying, outlier detection while abstraction and simplification are based on visual similarity. For instance, when classification is done by means of a fuzzy similarity measure, the degree of membership (uncertainity) can be employed into the visualization by adding further cartographic visual variables such as value and saturation to depict the original results of analysis

Data Transformation and Dimension Reduction (Projection):

When dealing with the visualization of high-dimensional information spaces it is difficult to make a choice among the different methods and their consequences. Phenomenon defined in high-dimensional information space is not directly accessible to human cognition. Although aim of dimension reduction is not only visualization, while visualizing multi-dimensional data they have to be transformed into low dimensional spaces while preserving relationships among information space components. Multidimensional scaling (MDS), principal component analysis (PCA), and self-organizing maps (SOM) are the most commonly used methods for reducing dimension to produce meaningful results even for large data sets.

For instance, losing a non-dominant dimension might cause loss of hot-spots. Awareness of these differences should influence the choice of a particular visualization method as well as the interpretation of its results.

Distance:

Maps, as the primary visualization geo-saptial data, typically attempt to preserve distance relationships where geographic analysis is also fundamentally founded on the First Law of Geography: "Everything is related to everything else, but closer things are more closely related" in evaluating geographic reality (Tobler 1970). The definition of "distance" might not be the same as well-defined geographic notion of distance. Modern geographic analysis utilizes a variety of functional distances such as travel time, Manhettan among others. Analogously, the choice of distance coefficients in information visualization is influenced by the goals of the representation as well as by the characteristics of the information space.

Generalization:

Despite the progress made in information visualization, there is still effort to increase limitations of graphic representation against scale, complexity and multi-dimensional nature of spatial data. In the field of cartography, tremendous expertise and knowledge is gained in dealing with complexities of representation of both model and graphic in a visual platform. The problem of graphic representation of the real world phenomena while resolving the conflict between number of features, size of visual signatures and available display media is an optimization problem. The conceptual and graphic meaning is preserved within context of geo-reality in spatial abstraction. If the process of abstraction and representation is scale-dependent, it is collectively named as cartographic generalization in mapping society, indeed it is geo-spatial abstraction in multi-granular graphic representation. The geo-spatial visualization and abstraction techniques experienced by cartography society for long time is relevant for visual spatial data exploration and visualization of consequences of spatial knowledge discovery process. For instance, classification and clustering in information processing is not different than aggregation and amalgamation in cartographic generalization respectively.

2.4 Spatial Data Mining

Various algorithms have been developed in disciplines such as statistics, machine learning, pattern recognition, database design, high performance computing, visualization and information theory to perform knowledge discovery (Kolatch 2001, Li and Yeh 2004, Miller & Han 2001).

Mathematical, analytical and statistical models are used for handling structured spatio-temporal tasks. However, ill-structured problems are handled by either heuristic or expert systems. Visual information processing and visual methods are used to understand problems (tasks), validate results when solving non-routine and ill-structured problems. Moreover, visual representations are useful for better understanding new concepts, constructing complex mental models, putting information into a generalized context, preventing information overload.

SDM is a demanding field since huge amounts of spatial data have been collected in various applications, ranging from remote sensing to GIS, cartography, environmental assessment and planning. SDM differs from non-spatial data mining because of the underlying spatial data where attributes of spatial objects are affected by attributes of spatial neighbours. In practice SDM comes down to a number of specific problem types and the techniques used to solve them (Clementini et.al.,2000, Han et.al. 2001-a, Han et.al. 2001-b). These are;

- Spatial Association Rules
- Classification
- Clustering
- Outlier detection

2.5 GIS: A Generic Spatial Data Handling Tool

GIS technology enables the user to visualize spatial data in the form of interactive maps, 3D models, dynamic visualization (animation), and multimedia. With its data collection, modelling, representation, management, integration, analytical and visualization potential for heterogeneous data (raster, vector, semi-structured, unstructured, stream data), GIS technology has proven itself providing support where spatial knowledge is needed (Longley et.al. 2003).

SDM whose objective is to discover interesting patterns in large spatial databases can be applied by using GIS tools. GIS software provides metric and topologic analysis functionality as a base to build tools for knowledge extraction, information discovery and decision making.

3. VISUAL SUPPORT FOR SPATIAL KNOWLEDGE EXPLORATION

Visual human interpretation is usually needed to create meaningful information after spatial analysis. Human ability of perception enables to analyze complex patterns *instantly*, recognize important patterns and make decision much more effectively (Foley 2000). Exploratory visualization comprises visualization of spatial data, consequences of SDM and visual data mining techniques. A series of visual data mining methods are connected to present the data in some visual form, to allow the human interactively get insight into the data by using visualization methods, to recognize patterns, to reveal relationships between data attributes, to identify global structures and finally draw instant conclusions directly from spatial data (Torun & Duzgun 2006).

3.1 Data:

Historical Data of Ship Accidents at Istanbul Bogazi (Istanbul Strait)

The Turkish Strait System includes the Straits of Istanbul, Çanakkale and the Marmara Sea, connecting the Black Sea and the Mediterranean Sea. The data about accidents has been collected since 1950's. More than 50.000 vessels, in average, annually use the Turkish Straits and this number has been increasing steadily. That means 15 passages out of 150 per day are tankers, half of which are 5% of which were tankers longer 200 m (one passage at every 10 minutes).

Accidents of shipping in the straits are examined under four categories: collision, grounding, fire and standing. When an accidents occurs accumulation of particles in the air and explosion during fire and heavy oil contamination formed on the surface of the sea and on the shores of Marmara Sea and the Istanbul Strait cause high risk of catastrophic disasters.

Legacy Crime Data In Ankara

The data of Ankara Police Directorate comprises type of crime incidents such as murder, burglary, auto, pickpocket and usurp with their address, location and time information occured in 2003.

Temporal Earthquake Data of Turkey

Istanbul, a mega-city is probably the most severe seismic hot spot world-wide with an estimated 0.41 probability of a severe earthquake, and possible tsunami, occurring over the next thirty years (Linnerooth-Bayer 2005, Erdik 2002). The earthquake data contains location of event, its magnitude and depth since 1900.

3.2 Visualizing Explored Spatio-Temporal Knowledge

Cartographic visualization allows user to interact with huge spatial datasets to recognize spatial distributions and, relationships and extract meaningful information by facilitating comparison, conceptualizing spatial patterns and processes, and other cognitive skills by making use of the visualization enabled system.

Visualization tools of GIS technology are used to model real world having different characteristics in the form of interactive, static and dynamic maps in 2D and 3D and at different levels of visual thinking (communication, synthesis, analysis and exploration). Human perception can be trained regarding different cartographic maps to develop a mental model of real world phenomena. Use of photo-realistic cartographic models and dynamic visualization might increase human perception to decrease the gap between real world and mental map (Fotheringgham et.al. 2000).

Several GIS systems have the capability of creating three dimensional perspective images using height or any other quantitative property of geographic features. This is used for illustration and visual exploration which may support cognition to design further analytical technics. To overcome other visualization problems such as bending and shadowing may be resolved by displaying from several different viewpoints.

Multi-dimensional characteristics of spatio-temporal data forces multi-variate and 3D visualization with combination of several techniques which support the main thema of the map. For instance, earthquakes having magnitude greater than 5 are visualized as proportional point symbol in third dimension on top of a reference map with fault zones (Figure 1). The map supports mental model about the location of devastating earthquakes along with fault zones and particularly the one parallel to northern shore of Turkey.

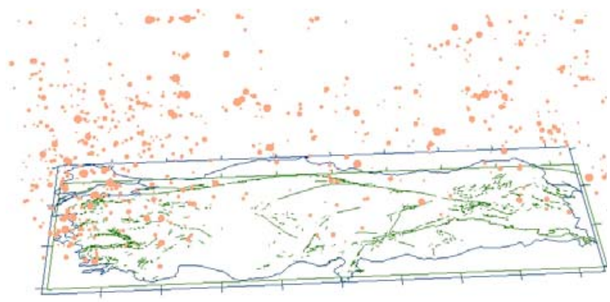

Figure 1: Location of earthquakes (Magnitude > 5), third dimension depicts time

Figure 2 depicts crime data on a landuse choropleth map. The crime type is represented with respect to association-selective perception property (hue). Color is one of the most effective tools to convey meaning in graphic representations. Time is shown in third dimension. Occurrence of a crime type with respect to land use and event time can be visualized in a single map which provides interaction, link plots and animation.

3.3 Knowledge Discovery by Using SDM Techniques and Their Visual Exploration

The increasing volume and diversity of digital geographic data easily overwhelm traditional spatial analysis techniques that handle only limited and homogeneous data sets with high-computational burden. The spatial data analysis and mining techniques can be grouped as generic, data specific, domain specific. The generic techniques have a statistical or strong

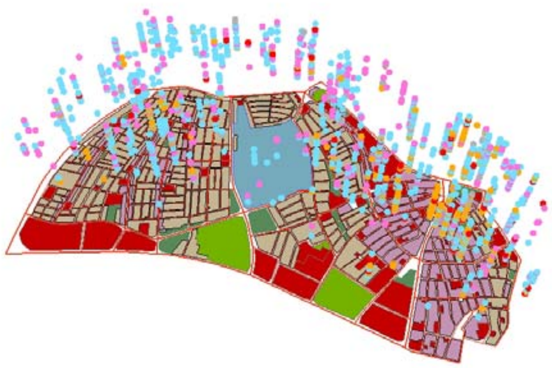

Figure 2: Crime locations, their types on top of landuse in Ankara

theoretical base. The data specific ones are usually regarding computational efficiency. The domain specific methods are mainly applied by a particular field.

Hot Spots (Getis-Ord): The G-statistic is often used to identify whether hot spots or cold spots exist based on so-called distance statistics. Hot spots can be detected by visualizing the distribution in format of choropleth or isarithmic maps (Baily et.al 1996, Shekhar et.al. 2003, Haining 2005, Getis & Ord 1996).

Local Autocorrelation (Anselin's Local Moran I) : Moran's I is a measure of global spatial autocorrelation. Global or local autocorrelation reveal feature similarity based location and attribute values to explore the pattern whether it is clustered, dispersed, or random (Longley et.al. 2003, Haining 2005).

In Figure 3, the accidents having autocorrelation with respect to location and overall accident index are depicted with darker colours indicating clustering. Clustering occurs at sharp turning points. The visualization is based on a multi-variant choropleth map where hot and cold spots are represented with red/green color contrast.

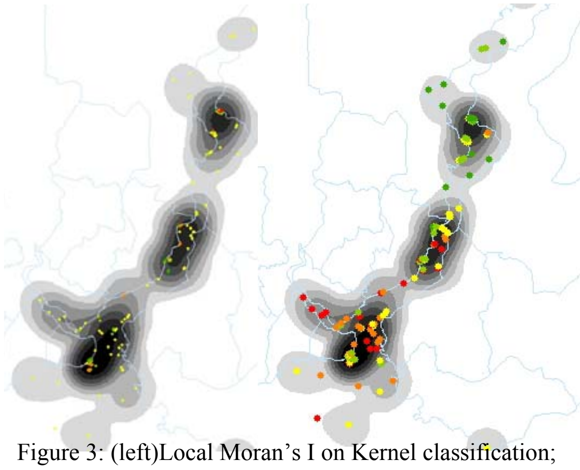

Figure 3: (left)Local Moran's I on Kernel classification; (right), Getis-Ord G statistic on Kernel classification

Density (Kernel) : Kernel density estimation is a nonparametric unsupervised learning procedure (classifier). Kernel is bivarate probability density function which is symmetric around the origin (Hastie et.al. 2001). The kernel analysis designates hot spots and distribution density of accidents respectively in Figure 3. Selecting kernel width causes a natural bias-variance trade-off as a change on width of the average window. The Kernel surface can be represented as any type of isarithm maps. Selection of Kernel width –level of detail- and its resultant distribution can be depicted either by means of animation which supports visual thinking and knowledge extraction on effect of spatial granularity.

ISODATA (Iterative Self-Organizing Data Analysis Techniques) algorithm : As being a variety of k-means, the ISODATA clustering method uses the minimum feature distance to form clusters to identify statistical patterns in the data. It begins with either arbitrary cluster means or means of an existing signature set, and each time the clustering repeats, the means of these clusters are shifted. The resulting values of centers and their covariance matrix are used to do supervised classification. In this step any of classification methods such as decision trees (dendograms) or ML is used (Mather 2004, Jensen 1996). Factors (census, cost of roads and distance to service locations) are partitioned into 5 clusters (Figure 4).

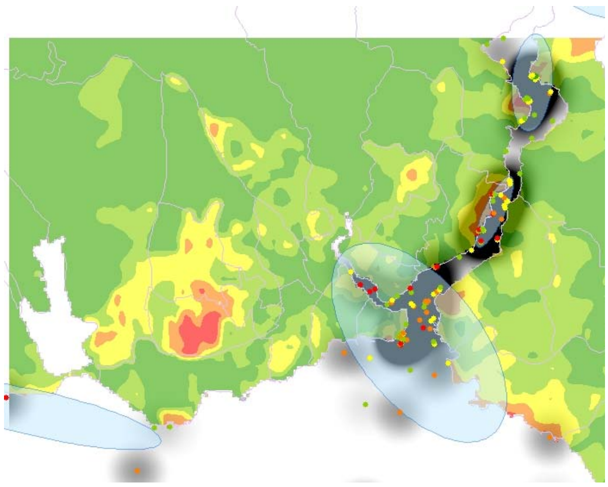

Figure 4: Multi-valued visualization: Vulnerability: Choropleth map; K-means clusters: Ellipses; Kernel density distribution: Raster Isopleth map; Acident location/severness: Colored dot map

Trends and Moving Mean Centre of Spatio-Temporal Phenomena: Mean of spatial data in linear time sequence (spatial-temporal moving average) is the moving mean center of a subset of observations drawn from a total sample which is listed in time order. Spatio-temporal moving average is used to detect changes in behavior and trend of the phenomena in time (Anselin 1992, Levine 2002). The use of visual hierarchies to express multi-level classification is not restricted to such hierarchical classifications based on thematic taxonomy. For example, one could simultaneously display k-nearest neighbors clustering solutions for a varying number, k of clusters.

The inherent hierarchy of the mapped features due to time is supported visually such as most recent trajectory are delineated with saturated colors and at higher-level in visual hierarchy.

Figure 5 depicts trajectory of moving spatio-temporal means and locations of ship accidents in a green-red color harmony where red is the closest in time. Red represents severeness of the accidents. The 3D graph shows that ship accidents have a trend to localize at mid sharp-turning point.

3.4 Evaluating Results

In Figure 4, it is clearly perceived that global kernel classification and temporal behavior of ship accidents are overlapping. The overlapping areas of vulnerability of people and frequent ship accident regions.

When the data is depicted due to day/night time, it is perceived that the most severe accidents have been happened in night time and it is localized at the most crowded part of the town namely Besiktas and Beykoz.

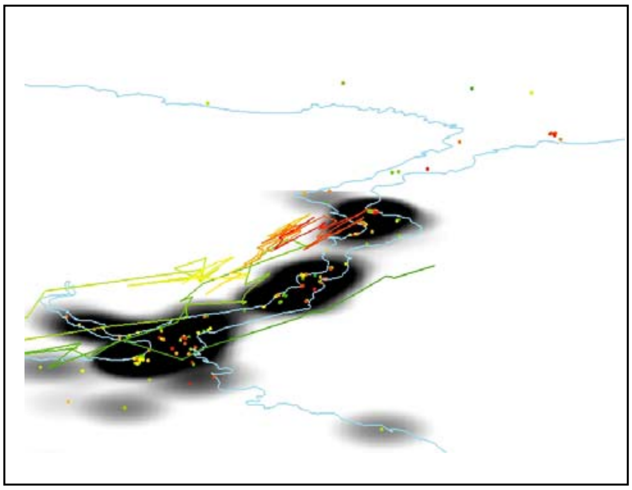

Figure 5: Trend of Spatio-Temporal Moving Average on Kernel density distribution

Three accident out of five locations are detected by using spatial data analysis (Local Moran's I, Getis-Ord statistic and Kernel density) and clustering (k-means). These are sharp turns of the Strait. Two regions in south are hot spot where severe accidents are populated.

The movement of spatio-temporal centroid of accident patterns from south to north might be a sign of extending trend of urban area through northern part of the strait by means of visual cognition. Due to dense population and civic centers around Istanbul Strait, vulnerability of people and places increases. This might cause severe disasters treating people, town, commerce and cultural heritage lying around the Strait.

4. CONCLUSIONS AND DISCUSSIONS

Evaluation methods to measure impact of visualization on knowledge discovery and effectiveness of visualization tools to increase cognition are further study directions. Although it is not studied, the movement of spatio-temporal centroid from couth to north might be a sign of extending trend of urban area through northern part of the strait. Changing lengths of ships might be another affect which prevents sharp turns. Due to dense population and civic centers around Istanbul Strait, vulnerability of people and places increases. This might cause severe disasters treating people, town, commerce and cultural heritage lying around the Strait.

A spatio-temporal cluster is an abstraction of moving clusters based on common spatio-temporal properties which are are being and having the same trend of behaviors such as being close in space, time and properties. The next step will cover determining behavior of clusters/classes in time that has more than one spatial dimension. Coupling data mining software such as Weka and Clementine to spatial analysis and visualization tools in GIS is going to be handled to use intermediate or final results in across platforms.

REFERENCES

Camossi,E. M. Bertolotto And E. Bertino(2006): A multigranular object-oriented framework supporting spatio-temporal granularity conversions, IJGIS Vol. 20, No. 5, May 2006, 511–534

Camossi,E. M. Bertolotto And E. Bertino(2006): A multigranular object-oriented framework supporting spatio-temporal granularity conversions, IJGIS Vol. 20, No. 5, May 2006, 511–534

Dzwinel,W., D.A.Yuen, K.Boryczko, Y.Ben-Zion, S.Yoshioka, T.Ito (2003): Cluster Analysis, Data-Mining, Multi-dimensional Visualization of Earthquakes over Space, Time and Feature Space, Earth and Planetary Sci. Letters, Aug.2003

Getis, A. and J . K.Ord (1996):Local spatial statistics: an overview. In P. Longley and M. Batty (eds), Spatial Analysis: Modelling in a GIS Environment. GeoInformation International: Cambridge, England, 261-277.

Gahagen M, M. Takatsuka, M. Wheeler, F. Hardisty (2002): Introducing GeoVISTA Studio: an integrated suite of visualization and computational methods for exploration and knowledge construction in geography, Computers, Environment and Urban Systems 26, 267-292

Han J., K. Koperski, and N. Stefanovic (1997): Geominer: A system prototype for spatial data mining. In Proc. 1997 ACM-SIGMOD Int'l Conf. on Management of Data(SIGMOD'97), Tucson, Arizona, 1997

Huang,Y. H.Xiong, S.Shekhar, J.Pei (2003): Mining Confident Colocation Rules without A Support Threshold, ACM 2003 SAC Melbourne, Florida, USA

Kötter, T. (2004): Risks and Opportunities of Urbanisation and Megacities, FIG Working Week 2004, Athens, Greece, May 22-27, 2004

Levine, Ned (2002).Crime Stat: A Spatial Statistics Program for the Analysis of Crime Incident Locations (version 2.0). Ned Levine & Associates, Houston, TX; National Institute of Justice, Washington, DC.

Li, X., Yeh A.G. (2004): Data mining of cellular automata's transition rules,Int. J. Geographical Information Science Vol. 18, No. 8, December 2004, 723–744

Koperski, K., J. Han (1995) : Discovery of Spatial Association Rules in Geographic Information Databases, Proc.4th Int'l Symp on Large Spatial Databases (SSD'95), pp. 47-66, Portland, Maine, Aug.1995

Malerba D., F.A. Lisi (2001_1): Discovering associations between spatial objects: An ilp application. In ILP '01: Proceedings of the 11th International Conference on Inductive Logic Programming, volume 2157 of Lecture Notes in Computer Science, pages 156–16, London, UK, 2001. Springer-Verlag.

Malerba D. , F.A. Lisi (2001_2): An ILP method for spatial association rule mining. In Working notes of the First Workshop on Multi-Relational Data Mining, pages 18–29. Albert Ludwigs Universitaet Freiburg, 2001.

Popelinsky L. (1998): Knowledge discovery in spatial data by means of ilp. In Zytkow J.M., Quafafou M.(Eds.): Proc. of 2nd European Symposium PKDD'98, volume 1510 of Lecture Notes in Computer Science, Nantes, France, 1998. Springer-Verlag.

Papadias, D., P. Kalnis, J.Zhang, Y.Tao (2001) : Efficient OLAP Operations in Spatial Data Warehouses, Technical Report HKUST-CS01-01

Shekhar S., Y. Huang (2001): Discovering Spatial Co-location Patterns: A Summary of Results, in proc. of 7th International Symposium on Spatial and Temporal Databases(SSTD01), L.A., CA, July 2001.

Tobler, W. A Computer Model Simulating Urban Growth in the Detroit Region. Economic Geography. 46(2): 234-240, 1970.

Torun, A., s. Duzgun (2006) : Using Spatial Data Mining Techniques to Reveal Vulnerability of People and Places Due to Oil Transportation and Accidents: A Case Study of Istanbul Strait, ISPRS Comm. II. Conference, GICON, Vienna, Austria

Zucker, J.-D., S. Mustiere and L. Saitta (2000). Learning Abstraction and Representation Knowledge: An Application to Cartographic Generalisation. The fifth International Workshop on Multistrategy Learning (MSL'2000), Guimarães, Portugal.

Anselin , Luc. (1992). SpaceStat: A Program for the Statistical Analysis of Spatial Data. Santa Barbara, CA: National Center for Geographic Information and Analysis, Universityof California.

Baily, C.B., A.C. Gatrell (1996): Interactive Spatial Data Analysis, Addison Wesley Longman

Chawla, S. et.al (2001): Geographical Data Mining and Knowledge Discovery, Eds.: Miller, H.J., J. Han, Taylor & Francis

Anselin , Luc. (1992). SpaceStat: A Program for the Statistical Analysis of Spatial Data. Santa Barbara, CA: National Center for Geographic Information and Analysis, University of California.

Baily, C.B., A.C. Gatrell (1996): Interactive Spatial Data Analysis, Addison Wesley Longman

Chawla, S. et.al (2001): Geographical Data Mining and Knowledge Discovery, Eds.: Miller, H.J., J. Han, Taylor & Francis

Cressie, Noel (1991). S ta tistics for S patial Data. New York: J. Wiley & Sons, Inc.

Fotheringgham, A.S., C. Brunsdon, M. Charlton (2000): Quantitative Geography: Perspectives on Spatial Data Analysis, SAGE Publications

Friedhoff, R., W. Benzon (1989). Visualization: The Second Computer Revolution. New York: Harry N., Abrams, Inc.

Haining, R. (2005): Spatial Data Analysis: Theory and Practice, Cambridge U. Press

Han, J., M.Kamber, A.K.H.Tung (2001-a): Spatial Clustering Methods in Data Mining, Geographical Data Mining and Knowledge Discovery, Eds. H.J.Miller, J.Han, Taylor and Francis

Han, J., M. Kamber (2001-b): Data Mining: Concepts and Techniques, Morgan Kaufman

Hastie, T., R.Tibshirani, J.Friedman (2001): The Elements of Statistical Learning: Data Mining, Inference and Prediction, Springer-Verlag

Longley, P.A, M.F. Goodchild, D.J.Maguire, D.W.Rhind (2003): Geographic Information Systems and Science, John Wiley & Sons

Miller, H. and Han (2001): J. Geographic Data Mining and Knowledge Discovery. Taylor and Francis, London, U.K.

Shekhar, S., Sanjay Chawla (2003): Spatial Databases: A tour, Prentice Hall

Erdik, M (2002): Earthquake Vulnerability of Buildings and a Mitigation Strategy: Case of Istanbul, IMM

Erdik M.(2005): Istanbul Master Plan, Bogazici University

Fernandez, J. (2005): Earthquake Master Plan for Istanbul, executive summary, Metropolitan Municipality of Istanbul, 2005

İstikbal, C. (2000): Oil Transportation Through The Turkish Straits:A Great Challenge To Safety, Turkish Maritime Pilots' Association (TUMPA), Annual meeting of MOIG (Mediterranean Oil Industries Group), İstanbul

Linnerooth-Bayer, J. (2005): Risk and Vulnerability Program, Research Plan 2006–2010

McCormick, B., T. DeFant, M. Brown (1987): Visualization. Scientific Computing in Computer Graphics: 21 i-E-8.

Munro, R., S.Chawla, P.Sun (2003):Complex Spatial Relationships, Tech.Rep.No. 539, ISBN 1 86487 584 4, School of Information Technologies, University of Sydney NSW 2006

Popelinsky L.,Jan Blatak (2005): Toward mining of spatiotemporal maximal frequent patterns, Masaryk University in Brno, Czech Republic

Clementini, E., P.D.Felice, K. Koperski (2000): Mining multiple-level spatial association rules for objects with a broad boundary, http://citeseer.nj.nec.com/cs

Foley, Jim. (2000): Getting There: The Ten Top Problems Left. http://www.computer.org/cga/articles/topten.htm

Kolatch , E. (2001): Clustering Algorithms for Spatial Databases: A Survey, http://citeseer.nj.nec.com/cs

Skupin, A. (): From Metaphor to Method: Cartographic Perspectives on Information Visualization

Department of Geography (www.geog.uno.edu/~askupin)
Spate, J., K. Gibert, M. Sanches-Marre, E.Frank, J. Comas, I. Athanasiadis, R. Letcher(2006): Data Mining as a Tool for Environmental Scientists, (accessed April 2006, www.iemss.org)