World Applied Sciences Journal 6 (Supplement 1): 46-53, 2009
ISSN 1818-4952

Extract of Landmarks for Pedestrian Route Guidance Systems by Means of Spatial Data Mining

A. Haghighat, A. Hasany, S. Fazeli, L. Eslami and S. Ghadiri

Department of computer, Islamic Azad University of Qazvin, Iran

Abstract: Since finding the route in unknown environments is difficult and challenging, great attention is given to solve this problem by means of automatic path finding calculation. The technology of route guidance systems has increased these days. As small and portable technologies are developing these days, a new group of users that is, pedestrians are also added to the previous users of route guidance systems. So new problems will occur in automatic guidance of computers because pedestrians are not connected to the routes network. As a result the use of route guidance systems for pedestrians is not practical instead; Landmarks_ preparing some recognizable and memorable guides through the route can be used to achieve this aim. The idea of this essay is to indicate; geographical data systems graphically which includes, spatial data, non-spatial data and neighborhood connection among them then we can use Spatial outliers to derive landmark from graph.

Key words: Spatial Data Mining · Landmark · Spatial Outlier · Route Guidance System

INTRODUCTION

With the increase of usage of geographical information systems great bases of various geographical data have been made available.Data searching on the data that have one or more special, spatial and or geographical is called spatial data mining and it is out put is the information and knowledge that have spatial and geographical feature like place, direction, distance, geometrical shape and the like. Geographical data searching is a kind of spatial data searching which considers objects or geographical spatial features.

Spatial knowledge findings show concepts about finding path by man that he needs note able and prominent object (or land marks) in order to find his path and direction in the space. But these objects are not basically landmark because being landmark is dependent property which depends on local milieu. Moreover human's ability in correct assessment of metric distances in week. Landmarks are objects with distinct environmental features that function as reference point. Landmarks help organizing the milieu and once it accompanies with navigating operation like turning to left or right considering that then and when this action should be done to make the navigating easy. Landmark can be divided into three categories: visual conceptual and structural. Absorption power of landmark is deter mind by visual absorption (like shaper color, visual space), semantic absorption (like historical or cultural significance, sings like stress signboards) and structural absorption (important junctions, rail or river, buildings).

For special group can be considered for land marks:

- Option point landmarks (at decision points)
- Landmarks having option potential(at point traveling)
- Landmarks on the way (a long a path without option)
- Land marks out of the way(far but visible on the way)

A study has shown that the option point and landmarks on the path are the most used in the unknown areas. So many researches have been done on spatial insight about the process of finding path and landmark theory but few of them due with the practical procedure and application for entreating automatic landmarks from available databases and from geographical information. The idea of the use of data mining for extraction landmark his been discussed in [1]. In this essay for the study of available spatial data base and automatically extraction of landmarks with the use of a process of knowledge discovering and using the techniques use of data mining have been shown. In problem [2] the extraction of landmarks in a virtual milieu has been discussed. Although there are deferens between airhead and real milieus, navigation techniques and finding the real world

Corresponding Author: A. Haghighat, Department of computer, Islamic Azad University of Qazvin, Iran

path here been exerted success fully to virtual milieus, Due to calculation limitations virtual milieus often have less visual, spatial and motor (moving) details. Outliners discovering algorithms were meaning full statistically and obtained outliers are simple recognizable by man. Thus the usage of there algorithms for extraction of landmarks from real miles can be very useful. In many works on the automatic extraction of landmarks. The collection of available spatial data has been used. More over in these works. As for as, I know just building for landmark option have been studied while recognizing researches on the subject of humans path finding with the help landmarks have shown that just buildings are not considered as landmarks. For instance a overpass traffic light river, park, are suitable landmarks. And can visually have so much attraction. In the second parh data input and in the third part the illustrations of geographical data geographically have been shown.

In the fourth part the spatial outlines witch one of the spatial data mining output have been studies and in the fifth part.

He Standard of appropriate land mark option has been stated and in the sixth part the land mark extraction with the help of spatial outlier discovering methods has bund studied. In the fifth the algorithm to spatial outlier discovering of mean and in the eighth can conclusion has been presented.

Input Data to Spatial Data Mining: Input data of spatial data mining are very complicated them input data to classical data mining, because they include extensive objectives like points, lines and polygons. Input data of spatial data mining have got two kinds to distinct features: non_ spatial features (for description to non_ spatial features of objects like name population and unemployment rate in a city) and spatial features (for defining spatial position and area and shape of spatial objects that often include information related to spatial position in space like length and with and height and shape. Communications among non _spatial objects in input data are obvious such as membership, subclass. Order, sample and versus communications among spatial objects are often implicit, such as mate covering, junctions and so on. Is a possible way for managing implicit spatial communications, converting communications to traditional input data columns and then the exertion to classical data mining techniques to them? Although giving real face to this communications (materialization) can result in losing of some information. Another way for obtaining implicit spatial communication, creating models or techniques for entering spatial information into spatial data mining process [3].

Graphically Illustration to Geographical Data: In the idea of the illustration, of geographical data has been geographically discussed this idea in for creating a model based on a graph for illustrating spatial and non _spatial.

Data and the use of this model for the production of the collection of data which in made of two kinds of data, in order to be able to exert a data mining technique on it and to consider the rich results of both kinds of data on objects and spatial relations between them.

In order to enrich the data mining process it in bather to consider all elements that has been used at geographical illustration (that is spatial objects, describing features and communications between them) these relations without doubt exrich spatial analysis process. For example we can understand the most important features of geometrical objects located at a particular point, or notice the patterns of representative hounds located with in the limits to high ways which pass through a province.

As we know one important feature of spatial data is that characteristics of neighbors of a particular object can have effects on the object itself.

Three types of spatial relations in this model is considered for data mining: Topological connections, orientation and distance. For stating the topological connections the model junction has been used. In this model topological connections between two things A and B in connection with internal object junctions A ($A°$), object limit A (∂A) and external object A(A^{-}) with internal abject B($B°$), object limit B (∂B) and extern object B (B^{-}) are defined outside of each object has been shown by its compliment. However this model can not make distinction between disconnected relations and can not distinguish topological relations between entities with holes.

$$\begin{pmatrix} A^0 \cap B^0 & A^0 \cap \partial B & A^0 \cap B^- \\ \partial A \cap B^0 & \partial A \cap \partial B^0 & \partial A \cap B^- \\ A^- \cap B^0 & A^- \cap \partial B & A^- \cap B^- \end{pmatrix}$$

The Figure 1 shows three disconnected relations between object A and B, however they have equal junntion_9 matrix because of elementary complements of object. In matrix their junctions have been that compliments cannot have a role in distinguishing disconnected relations.

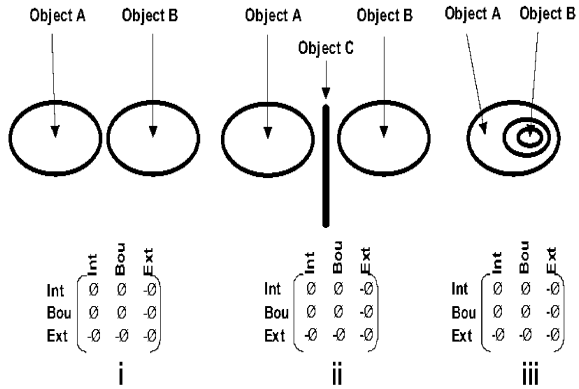

Fig. 1: Diverse disconnected relations with equal junction_9 matrix

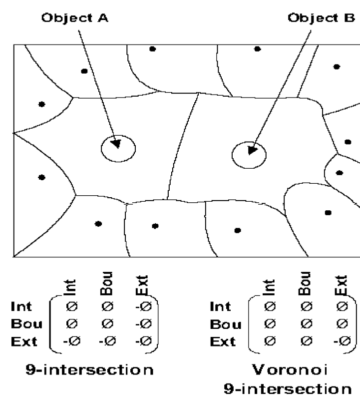

Fig. 2: Recognition disconnected relations

For the improvement of this condition model junction_9 based on Voronoi (V9I) as the corrected version to model junction _9 based on collection of points has been suggested this correction is exerted with the replacement of the external surface of a spatial object (complement) with its voronoi area. The voronoi of the first area is an entity and is defined with a spatial concept, the area with its effect and like an area including all places which are close to it than others. whites the help, of modal V9 the recognition of disconnected relation is impossible since each object instead of having connection with all other objects has limited neighbors. In the Figure 2 an example of result matrixes with the use of modal junction_9 and voronoi junction_9 from two disconnected objects has been presented. In the second state the external surface of a minus external surface B is the only connection which is not empty.

In the model based on graph, the spatial data is shown by outlines and manes. Outlines are used for the illustration of spatial objects and their features. The number of outlines in deter mined by the phrases

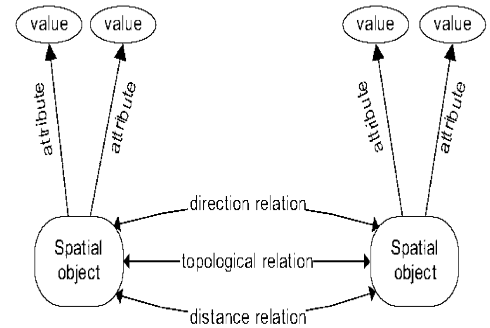

Fig. 3: Procdure suggested

$$n + \sum_{i=1}^{n} numAttributesPerObject(n_j)$$

Attributes per object (nj) n is the number of spatial objects within the given collection. The manes are shown the spatial connections between two particular objects. The capacity of this model for showing the connections within these objects has great effect on the result of data mining process. [5] The number of graph manes is deter mined by the following phrase.

$$\sum_{i=1}^{n} numAttributesPerObject(n_i) + numRelationsAmongObjects$$

The offered model has been shown in the Figure 3 in which two spatial objects are joined to each other through topographic connections, distance and direction.

These kinds of representing know ledye possess for creating graphs with the use of spatial and non _spatial data and spatial connections with in spatial objects as well.

Mixture of spatial and non _spatial data is one of the strengths of this model some the mining (searching) methods is exerted the data mining techniques on non _spatial data and then on spatial data and vice veresn it is offered that data mining techniques be exerted entirely on the collection composed of both types of spatial and non _spatial data. The Figure 4 in shown a database composed of ten objects and their spatial connections.

With offering a model based on graph a graph like Figure 5 produced when a graph was produced, it can be used as input data for a system of data mining based on graph [4].

Spatial Outlier: Outliers are in formally defined as observations in a data collection. Which is seen either they are incompatible with the rest to data collection or

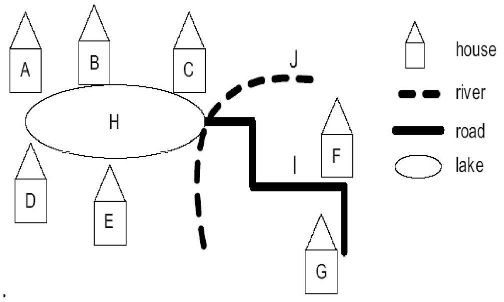

Fig. 4: Spatial data base object real world

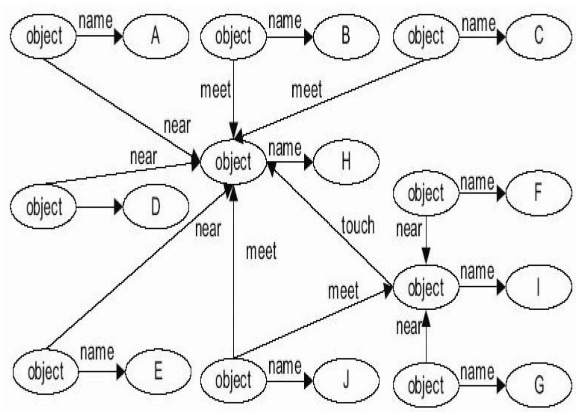

Fig. 5: Model based on graph in spatial relation

they are too different from other observations which make doubt that they are produced by other mechanism? Recognition of out liers is useful in many programs of geographical information systems and spatial data base including transportation. Zoology, public safey, public health, climate study, traffic study, services based on special position a … in general out liers are two general states: the entire out lier which can lend to the discovering of knowledge and unexpected patterns and it includes some applied programs in areas like swindling in credit card, the analysis to efficiency of athletes, disorder in voting and prediction of tough weather. It generally deals with numbers, characters and groups.

A spatial outliner is an object which is refereed to with spatial criteria and its non-spatial futures are meaning fully different from the rest of spatial object at one neighborhood. It processes very complicated data such as points, lines, polygons and three-dimensional objects and according to the first law of geography, spatial inter relation should be considered [2]. Method of discovering outlier can be classified into two vast groups with the names of one-dimensional methods of discovering out linear and multi-dimensional methods of discovering.

One-dimensional algorithms consider the mount of statistical spread of non-spatial features and ignore spatial relations between items. Methods of recognizing multi-dimensional outlier can be divided into two groups of heterogeneous methods based on metric and spatial methods. Heterogeneous multi-dimensional methods model data collection as a collection of points in a multi-dimensional is a metric space model and provide tests based on concepts such as distance and concentration. There methods have several limitations. First, they are designed for discovering the entire outliers than spatial outliers. Second, they assume that data items have been placed in a isometric space and they don't make a destination between spatial futures and non-spatial futures. Third, there methods don't extract inductive information about statistical spread data and last they seldom provide one safely criterion for discovered outliers.

Versions of spatial statistics provides two types of two part, multi-dimensional tests with the names of graphical tests and quantities tests for recognizing the outlier.

Graphical tests are based on illustrated spatial data and highlight spatial outliers. Among these methods change-script 2 and the dispersion diagram more on 3 can be mentioned.

For choosing spatial statistics, several parameters should be determined before the performance of test. First, neighborhood can be chosen based on a stable coordinating or a stable graphic distance or a stable Euclidean distance. Second, function of density of chosen neighborhood can be average, variance or the inter relationship itself. Third, for comparing a place with its neighbor's one digit or a vector of features, amount can be chosen. Finally, the statistics of base distribution can be chosen among various choices.

At [8] a method for discovering spatial outliers in data collection of structured graph known as Z has been presented. At this method definition of neighborhood in based on the connection of graph. For spatial statistics $S(x) =[F(x) - E_{y\in N(x)}(f(y))]$ has been used in which f(x) is the amount of features for a data record, N(x) is the stable cardinality of collection from the neighbors of X and $E_{y\in N(x)}(f(y))$ is the average of the amount of features for the neighbors of X.

S(x) determines the difference of the amount of features of each data X and the average of the amount of features of the neighbors of X. T outliers dB covering test is described as $Z_{s(x)} = \left|\frac{S(x)-\mu_s}{\sigma_s}\right| > \theta$

Is average amount of all S(x), σ_s is the deviation of all S(x). The option of θ depends on determined confidence distance.

The tests parameters are calculated by the algorithm called TPC. This algorithm with obtains a collection of data features and connected graph, obtains the knots of neighbors of each group of X and then calculates the amount of S(x).

[6] At first a general definition from S-outliner for spatial outlier has been suggestion and it gives amorality to entire spatial outlier. Then it describes the calculating structures of discovering methods of spatial outlier and presents flexible algorithms. The presented algorithm is similar to [8] and is design for discovering spatial outlier with the use of individual non-spatial features from data collection. This algorithm should be generalized for recognition of outliers with the use of several non-spatial features.

For this the definition of spatial neighborhood has not been changed but the functions of neighborhood density, comparison and statistical test should be corrected. This algorithm is divided into two groups: the repetitive algorithm 'r' and the repetitive algorithm 'Z' that both of them ignore real outliers and distinguish some incorrect outliers. This is considered as the advantage of this algorithm and in each occurrence one outliers is being recognized.

One method for discovering outliers has been offered at [7] that for calculating of being outliers of central object various weights are assigned to different neighbors. The weight is determined by spatial connection such as distance and the length of common border. The first offered algorithm is called the method weighed Z value which acts based calculation of Euclidean distance and discovering of outliers is done through the were of non-spatial features. For example the equation

$weight = \sum_{p=1}^{q} \alpha_p \bullet \frac{S_{pr}}{\sum_{i=1}^{k} S_{pi}}$ can be used for determining the weight. More than one spatial connection may cooperate in determining the weight.

The second offered algorithm at [7] is the averaged difference algorithm. This algorithm is based on the given average weight of averaged difference between X_i and each its neighbors. The main idea is that instead of obtaining the average of all neighbors of an object before comparison, object to be compared with each of its neighbors individually. In this algorithm spatial characteristics are used as difference weight between one object and its neighbors. In this algorithm there is no need to normalization since the difference is absolute.

Criteria for Choosing Suitable Land Mark: The main problem in using land mark in path-lending systems is distinguishing appropriate land marks and assessment of their suitability in operating path-leading instructions. So many researches have been done on determining suitable land marks criteria. The choice of suitable land mark depends on context, navigation concept and kind of application. In [9] several attraction criteria for choosing land mark have been presented. For visual attraction features of the area of external view of 6, shape, color, field of vision and other visual features have been considered. The area of external view of an object is an important feature for determining the contrast between the object and its circumstance. For example, the area of external view of a building with an ordinary shape (in the form of tetragon) is calculated with the multiplication of its height by the width. The shape of an object is determinant in its visual attraction. Unusual shapes against arbitrary objects of quite-tetragon attract the attention of people. Arbitrarily, the criterion for measuring an object is determined by considering the factor of its shape and deviation of its shape from the shape of tetragon.

The factor of the shape indicates the proportion of its height and width. The amount of difference derivation in the minimum area of tetragon limits the view of object as well as the area of object's view. One object according to its color can be more prominent than other surrounding objects. For Example a red building among a collection of gray building is more prominent the color of its object is usually determined by assigning the numerical amounts of the color chart of RGB. The feature of color is a problem in measuring and comparison.

The last visual feature which has been offered at [9] is the measuring of feature of visibility by calculation of the tow-dimensional field do vision. Visibility is being considered for the space which is used in movement (for pedestrians this space is the public street as well as some private ….). It is assumed that visibility is limited by recognize in which one pre-determined large area for any reason limits the visible space.

The amount and value of the field of vision of area of covered space by the core of the field of vision is derived in front of the object.

There are some other visual features for determining landmarks such as texture conditions ….. Of structure, which are not used be ……..of difficulty in recognition. For example the ax of the building is recognizable from the database but in the real world the building may seem new because of its outer reconstruction.

Other feature which is used for recognition of landmarks is semantic attraction. His tropical and cultural important of a place can cause is semantic attraction. For example, the square of ………….. In Tehran is known for its style of architecture. Information about semantic attraction can be obtained from database and available maps. At [9] one feature of Boolean is considered for each object and its being true means having historical or cultural importance.

Obvious signs like building's signs determine meanings for pedestrians. For example if a building has a restaurant or a museum it will have semantic attraction. This feature can also be defined as Boolean, that is a structure or a place with a distinct feature is being considered.

Structural attraction has an important role in the structure of spatial me lien, for example junctions and squares. Notes are among structural elements. For a driver a knot can be junctions and for pedestrians it can be locations. Structural feature of a ……..is its degree of connection. The degree may be given with quality of input and out put weight. Weights can be defined by a criterion of 5 (highways) to one pedestrian). Borders are other structural elements. Understanding structural features of borders depends on necessary energy for passing through them. For example lines of railroad disconnect two areas from each other and it can be passed only by passing over the bridge [9].

Extraction of Landmark with Help of Discovering Spatial Outlier: As it was already mention outliers are observation in a data collection which is dis compatible with the rest of that collection. Spatial outliers are objects that the amounts of their non-spatial features are meaning fully different from the amounts of spatial object of their neighborhood a landmarks can be considered as spatial outliers which their non spatial features such as height, color, volume, ... are different in comparison to their neighbors.

It this essay it is suggested that offered algorithms at [10] be used for discovering multi-variant spatial outliers from graph of offered geographical data at [4].

Regarding the detained measuring criterion at [9] for measuring the land marks absorption the vector of Y is defined as $y = (\alpha, \beta 1, \beta 2, \gamma \delta, \varepsilon, \zeta, \eta, \theta)$

Meaning of vector's features of constituents and an example of their amounts are given in Table 1. This way Q which is the indicator of number measuring criteria(features) equals Q. for each point of X in function the feature of f(x) is a vector Y.G(x) depending on definition of mean or median the amounts of features of Y on all points are at neighborhood of X(int real points NNK(x)). For simplification the amount of K in algorithm is considered fixed but K can be fined dynamically.

Difference between discovering methods of multi-dimensional outlier based on graph space and the outliers based on graph is through the definition of different neighborhood. Discovering methods of multi-dimensional outlier based on space use the distance for defining spatial neighborhood while discovering methods of outlier based on graph use the graph connection. There have the algorithm of spatial outlier discovering can be changed for recognizing the outliers that are produced in graph GIS.

Table 1: Properties landmarks

Properties for visual at-traction	Example	Measurement
Façade area	α = 25m * 15m = 375sqm	$\alpha = \int x \mid x \in facade$
Shape		
• Shape factor	β_1 = 15m / 25m = 0.6	β_1 = height / width
• Shape deviation from rectangle	β_1 = 375sqm − 295sqm = 21%	β_1 = (area of minimum bounding rectangle − α) / area of minimum bound-ing retangle
Color	γ = [255, 0, 0] = red	γ = [R. G. B]
Visibility	δ = 2400 sqm	$\delta = \sum x \mid y\ visible$
Cultural and historical importance	ε = T	$\varepsilon \in \{T. F\}$
	ε = 1	$\varepsilon \in \{1,2,3,4,5\}$
	(Building very famous for its architecture)	Scale of importance: 1(high) − 5(low)
Explicit mark	$\zeta = T$	$\zeta \in \{T. F\}$
	Sign on front of a building	Boolean
Nodes	η = (4*2+4*2)	$\eta = (i+0)$
	Second node in Figure 2 for pedestrians: all streets are town streets (w=2)	weighted incoming (i) and outgoing (o) edges to and from a node
Boundaries	θ = 2500	θ = cell size * form factor
	Channel dividing a district	From factor: long side / short side

Neighborhood at [1] is defined as circles with radius of 50 meters and 100 meters around the decision point. At [8] the depth of neighborhood in graph is received as input parameter. Depending on the collection of input spatial data (data base or graph) this neighborhood definition can be used. More research should be done on the detection of suitable neighborhood. The movement direction of a pedestrian and his field of vision should be considered in defining neighborhood. An object should be chosen as land mark that in front of which there is the pedestrian and he should be able to recognize it be for stooping at the decision point.

Algorithm of Discovering the Spatial Outlier of Mean: Algorithm of recognizing spatial outlier of mean is suggested at [10]. This Algorithm as already mentioned is presented for discovering multi variant spatial outliers. This Algorithm with obtain a collection of spatial points, a feature (specialty) vector, a neighborhood vector and a comparison vector, finds spatial outliers.

A spatial graph G (D, E) composed data collection $X = \{x_1, x_2, \ldots, x_n\}$ pre-determined defined threshold ofθ, specialty vector F and number of K is taken from the closet neighborhood. For each fixed $1 \leq j \leq q$, spatial (feature) vector of $f_j(x_i) \leftarrow \frac{f_j(x_i) - \mu_{fj}}{\sigma_{fj}}$ becomes standard, that is i=1, 2, 3,...

For each spatial point the Xi collection of K the closet neighborhood NNK (Xi) or the collection of K notes of neighborhood of Xi in graph G, is calculated.

For each spatial point the Xi neighborhood vector of G is calculated as gi(Xi) is equal to the average of data collection.

$$\{f_j(x): x \in NN_k(x_i)\}$$

And the comparison vector be

$$h(x_j) = f(x_i) - g(x_i)$$

The following amounts are calculated as well.

$$d^2(x_i) = (h(x_i) - \mu_s)^T \textstyle\sum_s^{-1} (h(x_i) - \mu_s)$$

$$\mu_s = \frac{1}{n}\sum_{i=1}^{n} h(x_i)$$

$$\textstyle\sum_s = \frac{1}{n-1}\sum_{i=1}^{n} [h(x_i) - \mu_s][h(xi) - \mu_s]^T$$

If $d_2(x_i) > \theta$ Xi is a spatial outlier. Outliers are added to a collection so finally an outlier with the biggest amount of h(x) is chosen as a land mark. Experiments have shown that this method is practically elective and it can correctly find spatial outliers.

CONCLUSION

Nowadays systems of navigation and path direction of pedestrians have been extended. Landmarks play an important role in path finding process. Landmarks are prominent and distinct objects which are easily distinguished by the use discovering techniques of spatial outliers which is one the most important patterns pf output spatial data mining, for extraction out land marks from system of geographical in for motion.

So geographical data can be changed to a graph the shows both spatial data and non-spatial data and spatial relations between data. This graph can be used as algorithm's input of data mining.

Outlier discovering techniques were meaning for statistically and their efficiency has been proved. So the use of these techniques for extraction landmarks should be done on real data. Recognizably of extracted land marks should be surveyed by the users and pedestrians. One of the problems in this case is that in different conditions the quality of land marks are different. For example during the night or in a rain and foggy weather path finding with the land marks which are used along the day is different.

On the other hand, the complexity of time and calculation of algorithms of spatial data mining is of the high, but it needs high speeds in path guidance system. So one of the coming works which has stated in most essays is decrees of complexity and increase of speed for application of spatial data mining algorithms. Research on temporary spatial patterns is also one of the issues which have been considered in presented essays.

REFERENCES

1. Elias, B., 2003. Extracting landmarks with data mining methods, W. Kuhn, M. Worboys and S.Timpf (Eds.), Spatial Information Theory: Foundations of Geographic Information Science,Lecture Notes in Computer Science 2825, Springer-Verlag, pp. 375-389.
2. Lazem, S.Y., W.M. Sheta and M. Kantardzic, July 2005. Automatic landmark identification in large virtual environment: a spatial data mining approach, Proceedings of the Ninth International Conference on Information Visualisation (IV'05)æ IJCIS, Vol 5, No. 1.

3. Shekhar, S., P. Zhang, Y. Huang and R. Vatsavai, 2004. Trends in Spatial Data Mining, as a book chapter in Data Mining: Next Generation Challenges and Future Directions. H. Kargupta, A. Joshi, K. Sivakumar and Y. Yesha (editors), AAAI/MIT Press.
4. Palacio, M.P., D. Sol and J. González, 2003. GRAPH-BASED KNOWLEDGE REPRESENTATION FOR GIS DATA, Proceedings of the Fourth Mexican International Conference on Computer Science (ENC'03).
5. Shekhar, S. and S. Chawla, 2003. Introduction to Spatial Data Mining, in Spatial Databases: A tour, Prentice Hall, ISBN, 013-017480-7.
6. Shekhar, S., C. Lu and P. Zhang, 2003. A unified approach to detecting spatial outliers, GeoInformatica, 7(2): 139-166.
7. Kou, Y., C.T. Lu and D. Chen, 2006. Spatial Weighted Outlier Detection, Proceedings of the Sixth SIAM International Conference on Data Mining, April 20-22, 2006, Bethesda, MD, USA. SIAM.
8. Shekhar, S., C.T. Lu and P. Zhang, 2001. Detecting graph-based spatial outliers: algorithms and applications. In Proc. of the 7th International Conference on KDD.
9. Raubal, M. and S. Winter, 2002. Enriching wayfinding instructions with local landmarks, In M.Egenhofer and D. Mark (eds), Geographic Information Science, Lecture Notes in Computer Sci., 2478, Springer Verlag, pp: 243-259.
10. Lu, C.T., D. Chen and Y. Kou, 2003. Detecting Spatial Outliers with Multiple Attribute, IEEE International Conference on Tools with Artificial Intelligence.