

Visualization-Driven Approach to Anomaly Detection in the Movement of Critical Infrastructure

Evgenia Novikova^{1,2(✉)} and Ivan Murenin¹

¹ Department of Computer Science and Engineering,
Saint Petersburg Electrotechnical University “LETI”, Professora Popova Street 5,
Saint-Petersburg, Russia

imurenin@gmail.com, novikova.evgenia123@gmail.com

² Laboratory of Computer Security Problems, St. Petersburg Institute for Information
and Automation (SPIIRAS), 14th Line, 39, Saint-Petersburg, Russia

Abstract. Detection of anomalies in employees’ movement represents an area of considerable interest for cyber-physical security applications. In the paper the visual analytics approach to detection of the spatiotemporal patterns and anomalies in organization staff movement is proposed. The key elements of the approach are interactive self-organizing maps used to detect groups of employees with similar behavior and heat map applied to detect anomalies. They are supported by a set of the interactive interconnected visual models aimed to present spatial and temporal route patterns. We demonstrate our approach with an application to the VAST MiniChallenge-2 2016 data set, which describes movement of the employees within organization building.

Keywords: Cyber-physical security · Spatiotemporal movement patterns · Anomaly detection · Self organizing map · Heat map

1 Introduction

A wide range of applications may benefit from analysis of moving point data sets, also known as trajectories. Mining trajectories helps to establish object life patterns, reveal constraints existing in the underlying environment, for example, rules or security policies, restricting access to the specified zones, or infrastructure available to the individuals, i.e. places of interests, ATMs, etc. [1]. Another important application of the trajectory analysis is the generation of the features of the object life pattern to detect possible anomalies in observing data sets [2]. The roots of the detected anomalies may be different: they can be rather harmless like a driver encountering problems in the unknown place, or they can be signs of possible crime, for example an employee violating company security policies. Thus, the analysis of the trajectories can be used to provide air, road traffic and maritime safety by monitoring location, speed, trajectory and detecting unexpected travel impediments. Monitoring movements of the employees of critical infrastructure and hazardous industries supports monitoring compliance of the safety and access control policies and is useful in detection of inside threat [3].

This paper presents an approach to detection of the spatiotemporal anomalies in organization staff movement based on a combination of data mining and visualization techniques. A set of interactive visualization models supports understanding of the existing patterns in employees' movements and highlights possibly anomalous situations. We propose to form behavior patterns depending on a week day as daily routine of the company staff may vary depending the week day. To understand what temporal patterns in employee's movements exist, we develop a special glyph that displays a set of days of week sharing the same pattern. Patterns are displayed using two visual models – the one is stacking based model with time bar reflecting the sequence of the monitored zones visited by employee, and the second one is a graph of connected controlled zones used to reflect spatial attributes of the pattern. The deviation in employees' movements is displayed using a heat map linked to detailed view on moves. Thus, an analyst has a possibility to understand the character of the detected anomaly.

Specifically, our main contribution is an approach to analysis of the movements of critical infrastructure stuff that handles multi-dimensional data (including employee groups and job classification) with the temporal features for the context of physical security. Its key elements are the interactive self-organizing maps used to detect groups of employees and days with similar behavior, enriched with specially designed glyph able to reflect periodicity in movement depending on a day of week and a heat map applied to detect anomalies and enforced by anomaly ranking mechanism.

To demonstrate the efficiency of the proposed approach we tested it on data set provided within VAST Challenge 2016 [4].

The rest of the paper is organized as follows. Section 2 discusses the related work on approaches to anomaly detection in time series and visualization models used to investigate object movement. In Sect. 3 we describe the proposed visual analytics approach to anomaly detection in employee movements. Section 4 presents case study used to evaluate approach, discusses results and defines directions of the future research. Conclusions sum up our contributions.

2 Related Work

In case when there is no possibility to obtain patterns of normal behavior by making observation the most widely used approach is based on the clustering of the trajectories. The obtained clusters are then used to describe normality model for anomaly detection. Clusters may be found by centroid based approaches, hierarchical models, or density-based approaches [5]. Automatic methods may discover interesting behavioral patterns and anomalies but in the most cases they need to be supported by the visualization techniques explaining the final result [6, 7]. A neural clustering method, also known the self-organizing map (SOM) combines multidimensional data clustering and projection techniques, and produces visualization of the clusters reflecting distance between them. Shreck et al. applied SOM to analyze trajectories and propose a visualization-driven framework for adjusting SOM output [8]. However, they focused mainly on the analysis of spatial attributes of trajectories. In [9] both spatial and temporal attributes are investigated; however, they are analyzed separately by selecting all temporal attributes for

one geographical location or all spatial attributes for one time unit. Authors enriched SOM visualization by special images displayed within SOM cells. The goal of these images is to explain the result of clustering tool.

In general case existing models could be divided into three groups – static or interactive maps often enriched with glyphs encoding movement attributes, space-time cubes, and stacking based visualizations [6].

Maps are the most obvious way for presenting location aware data. Trajectories or cluster of trajectories are represented by lines. The movement attributes such as time, speed, type of the moving object are encoded by line color or specially designed glyphs. For example, in [10] the color of lines presenting routes of moving points is used to encode type of vehicles (car, bus, pedestrian, bicycle and others) and speed values. In cases when the exact trajectory is not important flow maps are used. Flow maps are visualization models focusing on determining destinations and sources of the routes. The quantities of the flows are usually mapped to two visual variables: the width and the color saturation of the flow lines, and the glyphs are used to characterize the type of the flow destination or source. Interesting modification of the flow maps used to analyze population migration is presented in [11]. The regions of interest serving as destination and source points of migration are located circumferentially and displayed as segments of the ring. The flows are displayed as splines connecting corresponding ring segments. However, all map-based visualization share one common disadvantage – they are not able to display spatial and temporal attributes of the movement simultaneously.

Space-time cube is a 3D-visualization technique designed to present spatial and temporal characteristics of the movement simultaneously. According to it, points of the trajectories are displayed in three dimension space, where vertical axis stands for time usually. In [12] authors presents an extension of the space-time cube called trajectory wall. In contrast to traditional space time cube the third display dimension is used to represent a set of trajectories. The vertical axis is divided into bins, each containing one trajectory. Trajectories are represented by bands split into segments, which are colored according to the values of attributes related to trajectory points. As the trajectories can be ordered in the third dimension according to their temporal order then this visualization model can be viewed as a space–time cube where the absolute time is transformed to the temporal order of the trajectories. Like all 3D visual models the space time cubes could be ineffective because of occlusions and cluttering of the trajectories.

Stacking based visualization of the routes is based on time graph, also known as time line. One axis represents time, and another – values of spatial attributes of the trajectory point. Each route is represented by a curve or polygonal line. To display a set of trajectories, stacked curves or bands synchronized within time scale are used. Bands can be divided into segments colored according to values of attributes. For example, in [13] color is used to display type of position (café, shop, office).

3 Visualization Driven Approach

When designing our approach to analysis of the employees' movement we tried to answer on the following questions the analyst would be interested in:

employees, employee etc. in table view. All views are interactive and interconnected. Clicking on each graphical element of all data visualization models an analyst updates information displayed in the linked views. In the subsections below we discuss the data preprocessing step and describe proposed visualization and interaction techniques in detail.

3.1 Data

In general case logs from proximity card readers describing employees' movement have following format: <timestamp, employee ID, controlled zone ID>. There are could be some additional fields such as status (entrance permitted or denied), employee access level, etc. These data could be enforced by description of the control zone arrangement, employees' position within organization hierarchy, location of the employees' work place in the context of the controlled zones. In our approach we deal with logs containing information only who and when entered controlled zone, and show that these data are enough to form behavior patterns and discover signs of the anomalous activity.

The specific feature of the logs from proximity card readers registering employees' movement is that they appear irregular, and interval between logs for one employee may vary from tens of seconds to tens of hours. Furthermore, some employees can make a lot of moves within organization building due to their role profiles, while others rarely leave their work place. Thus, the lengths of time series describing their movement may vary significantly.

In our approach we transform each time series to a vector of the fixed length. We divide the whole time interval presented in logs into a sequence of the equal time slots, and calculate number of visits and duration of staying in each controlled zone for each time slot and each employee. The attribute of the vector generated in this way describes an activity of the particular user in the given controlled zone during given period of time. The attributes are ordered by time slots firstly and then by controlled zones. Currently the default duration of the time slot is 4 h, experiments showed that it is enough to detect even minor temporal deviations in movement equal to 5 min.

3.2 The SOM-Based Views

To detect groups of employees with similar behavior and outliers, we use SOMs known also as Kohonen maps [14]. It is a type of artificial neural network that is trained using unsupervised learning to map multidimensional input data into a low-dimensional (typically two-dimensional) space. The SOM is a topology-preserving data transformation technique meaning that the cluster centers associated with nodes located next to each other are more similar than clusters located far from each other.

One of the major problems with SOMs is obtaining data without missing values for each attribute. However, in our case data preprocessing step guarantees producing vectors with values for each dimension.

In our approach we use U-Matrix presentation of the SOM [15]. It shows data structure by displaying the average distances between weight vectors of neighboring units. The darker color of the node, the more it differs from the neighbors. The adjacent light

nodes are quite similar to each other. Nodes containing clusters' centers are marked with circle glyph. Its size reflects the number of objects in the cluster.

The SOM is used twice in the proposed approach. The purpose of the Employees SOM view is to detect groups of co-workers with similar behavior. The attribute vector describes activity of the employee during all period of time, and deviations in the movement taking place once or rarely do not influence on the result of the clustering. This enables us to assume that this SOM shows differences in movement existing due to peculiarities of the employee's role in the organization. Each element of the SOM view is selectable; by clicking on it an analyst gets detailed information about cluster members, and updates the view of the second SOM, the Graph View, the Pattern View and the Heatmap View.

We assume that employees may have responsibilities implemented on some regular basis depending on the day of week; these duties may cause periodical changes in the employee's movement. The goal of the second SOM is to detect groups of days with similar behavior for the selected group of employees revealing thus periodicity in their movement. To construct the Days SOM, we use weight vectors of cluster centroids of the Employees SOM, transform it to a set of vectors representing days by splitting it into vectors of smaller length and cluster it using SOM. Then we determine what day of week contains in each group of days to discover dependencies of the movement from day of week. The result of clustering is displayed using U-Matrix, however, the SOM nodes are complemented with a special glyph that displays the distribution of the days in the cluster according to the day of week. In many organizations the employee's activity routine depends on type of the week in the year – odd or even. We designed a *WeekCircle* glyph able to display movement patterns having periods equal to one or two weeks. It may be divided into 7 or 14 sectors depending what periodicity model – 7 day or 14- day is selected. The model of the glyph for the 14-days period is shown in the Fig. 2. The right half of the WeekCircle represents odd week, and the left half of it represents even week. The Mondays are displayed by the top sectors and Saturdays – by the bottom sectors, thus odd week is in mirror reflection with even one. Figure 2 presents two glyphs reflecting what days of week the group contains. The left glyph shows that the cluster consists of Mondays, Wednesdays and Fridays of the odd week meaning that the employee or group of employees has particular duties implemented every second Monday, Wednesday and Friday. The right glyph indicates that the group contains weekends only. The Days SOM view allows also detecting days with anomalous behavior if these anomalies have long term character, i.e. their duration exceeds an hour. The day with such type of anomaly would constitute a separate cluster located in the close neighborhood to another cluster.

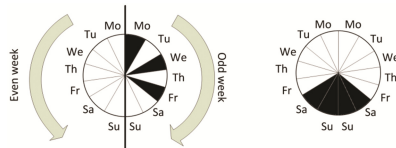


Fig. 2. The *WeekCircle* glyph showing the distribution of the days according to the day of the week

We implemented interaction techniques for the Days SOM view similar to the Employee SOM view. An analyst can set up the SOM size, get detailed information about each SOM cluster. By clicking on the node of the SOM the analyst updates the Property View, the Graph View, the Pattern View and Anomaly View.

3.3 The BandView Visualization Model

The goal of the BandView model is to link spatial and temporal attributes. It is a stacking based visualization technique. The horizontal axis corresponds to the time. The route of an employee is presented using segmented bar. Each segment represents time interval during which the employee was in the given controlled zone. The color of the segment is used to encode the zone itself. The color scheme is constructed in the following way. Each floor of the organization building is assigned a certain color. The palette for the zones located on one floor is created by changing brightness of the floor color. The greater the number (ID) of the controlled zone, the darker the color of the corresponding segment. The Fig. 3 shows routes of the co-workers of one department during one day. The zones of the first floor are displayed in brownish colors, zones of the second floor – in greenish colors, and zones of the third floor – in blue colors. Obviously, the majority of the employees spend their work time on the second floor, and only one person has an office on the third floor.

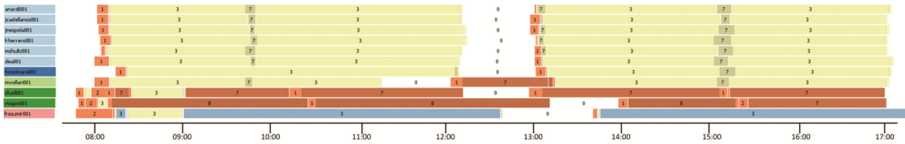


Fig. 3. The BandView representation of the employees’ moves during one day.

The BandView model is used in the Pattern View and the Anomaly View. In the Pattern View it displays spatiotemporal patterns for the groups of employees and is used in conjunction with the WeekCircle model. The Y-axis of the BandView represents clusters of the days with similar routes. If no node in SOM-based views is selected the Pattern View displays raw data. In this case Y-axis represents staff members ordered by the department. To analyze raw data, we implemented flexible filtering mechanism allowing constructing complex logical expressions using all available attributes of the move: employee ID, department, office, duration of staying in the zone, zone ID, floor, etc.

In the Anomaly view it shows movements of co-workers belonging to one cluster during one particular time interval (group mode) or routes of the employee during days belonging to one cluster (day mode). An analyst may switch between these two modes, however their availability is determined by the structure of the employees’ cluster. If it consists of more than two persons then both modes are available to the researcher, otherwise the day mode is used. In the group mode Y-axis represents employees forming one cluster, and the scale of the timeline is limited to one day. In the second mode Y-axis displays a set of days belonging to one cluster ordered by time. However, in both

modes BandView allows detecting where, when anomaly took place and how long it lasted; because it enables visual comparing of the routes.

3.4 The Graph View

The graph is used to display controlled zones visited by employees. The graph vertexes corresponds to controlled zones, adjacent zones are linked by edges. The graph of the controlled zones can be constructed on the basis of logs from proximity card readers or map of controlled zones.

The Graph View is controlled by two views – the Employees SOM View and the Days SOM View. These two views determine what data are to be displayed. When clicking on the element of the Employees SOM, the Graph View displays zones visited by a group of employees. The Days SOM View refines data to be displayed – the Graph View shows spatial pattern of the movement for this group during selected set of days.

The zones visited by the employee are mapped on this graph and highlighted by the color. They are colored in accordance with color scheme used in the BandView visualization model. The unvisited zones are displayed in grey.

3.5 HeatMap View

The goal of the heat map is to show the presence of possibly anomalous deviations in the staff personnel movement. We consider that anomalies come out in irregular insignificant changes in the behavior of the subject, therefore we suggest investigating the deviations within group of employees or group of days having similar movement pattern. Like the Anomaly View the Heatmap View has two modes. In the first mode it displays deviations in routes of employees constituting one cluster. The heat map is constructed in the following way. The Y-axis corresponds to the employees in the cluster; and X-axis represents attributes of the vector generated from the log data as it is described above. In the second mode Y-axis displays days of one cluster, and X-axis displays attributes of the vectors describing person activity during given day. Each element of the heat map represents distance of an attribute value of the sample from the cluster centroid.

However, displaying the distances directly may produce rather noisy picture on the one hand when the distances are comparable, or hide some deviations on the other hand, if distance variance is high. To solve this problem we use anomaly ranking mechanism based on calculation of the z-score for each attribute deviation from cluster centroid. Z-score reflects how far the current value of the attribute from its mean value. Discretization approach used on the data preprocessing stage of the approach allows forming data samples for the time slots characterizing employee activity for the given period of time considering some time periodicity (day, week, month, etc.). This makes it possible to assess deviations in employee route, for example, on the time interval from 8 am till 12 pm in the morning for each work day.

The zones of the heat map with suspicious bursts of the activity can be selected and examined in detail in the Anomaly View.

4 Case Study and Discussion

To evaluate our approach we use a dataset provided by the VAST Challenge 2016: Mini-Challenge 2 [4]. It contains logs of the proximity card readers that cover individual building zones. When an employee with proximity card enters a new controlled zone, his/her card is detected and recorded. It should be noted that most, but not all, areas are available to staff members even if they forget their proximity cards. The dataset contains a two-week set of logs. An analyst is also provided with building layout for the offices, including the maps of the controlled zones, a list of employees, including their department and office assignments.

We assumed that the employees within one department may have similar movement patterns and decided to analyze logs grouped by the employee department. We discovered that the majority of the employees within one department moves alike, and there are small groups whose behavior is different due to the specific responsibilities and location of the work place. Figure 4 shows the result of clustering of the employees of the security department. It is clearly seen that there are 5 groups of co-workers having similar movement pattern. One of these groups is rather numerous, while others consist of one-two persons.

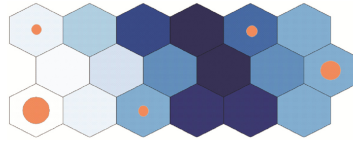


Fig. 4. Result of the SOM clustering of the employees belonging to one department

The members of the most numerous group have only two periodical patterns in their movement – one is for the work days and another - for the weekdays. Their work days usually start approximately at 7.50 am and finish at 5 pm. They leave the building approximately at 12 pm for an hour. We consider that they go outside for lunch. The members of this group spend the most of their work time in the 2–3 zone located on the second floor where their offices are located. Every 1, 5–2 h they leave zone and visit adjacent zone 2–7 for 5–10 min. The 2–7 zone contains offices and toilets. As they visit only one zone of the second floor this route hardly could be considered as a go-round, more likely they simply go out to refresh. At the weekends they do not come to work.

Analysis of the heat map displaying deviations in their routes allowed us to spot easily anomalies in the routes of two employees on the second day of the period investigated (Fig. 5). The BandView visualization of the movements of the selected group of employees for this day showed that the logs from proximity card readers are missing for one employee while another employee has doubled logs. Interestingly that the second doubled log has a timestamp several seconds later than the first one. These two facts – the absence of the logs for one employee and doubled logs for another – allowed us to assume that the first employee used the proximity card of the second employee to enter the controlled zone.

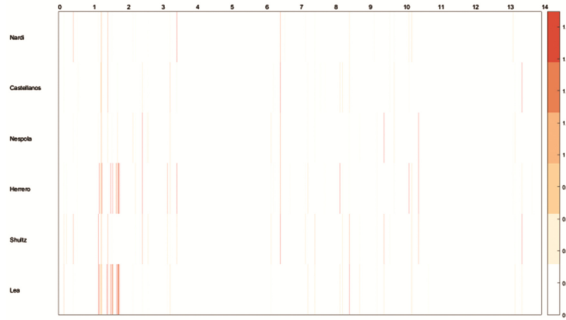


Fig. 5. Deviations in the routes of the most numerous group of the security employees

Two clusters closest to this group consist of one employee each. Their routes are rather similar to movements of the employees of the first cluster as their offices are located in the 2–3 zone, where they also spend most of their working time. However, one employee (the upmost left cluster) does not leave 2–3 zone in the first half of the day. The distinctions in the route of employee presented by the right cluster in the bottom row of the SOM are much greater – he leaves building for lunch one hour earlier at 11 am in order to be in the 1–7 zone during standard lunch time.

The rest two groups lay apart from the others in the SOM. The rightmost cluster consists of two employees whose offices are located in the first floor. They expose two periodical patterns depending on the day of week. One of them visits the 2–3 zone every Tuesday in the first half of the work day, while another visits the same zone every Thursday in the first half of the day. We spot interesting deviation in moves of the one employee of the given cluster. He visited 3–4 zone for 3 min located on the third floor only once during all two-week period. According to the zone plan lifts and stairs are located there. We could assume that the employee visited this zone accidentally, for example by pressing wrong lift button. However, to be more precise we need more information about who else from the staff was in this zone at that moment. The last cluster located in the upper right corner of the Employee SOM contains only one employee. From the BandView it is clear that his/her office is in the third floor. Analysis of the office assignments shows that the third floor is occupied by representatives of the administrative and executive department, thus allowing us to admit that this employee is a chief of the security department. His moves are rather diverse and strongly depend on the day of week. He starts his work day with visit of the local café located in the 1–2 zone and then goes upstairs in his office. Every Tuesday and Thursday he goes to the 2–3 zone and spends there approximately half an hour and then returns to his office. The BandView of the raw data for the moves of the security staff allows us to conclude that the department meeting takes place every Tuesday and Thursday. It is also possible to find out that the representative of the security staff has to be on the first floor from 8 am to 5 pm. For this reason some security staff has lunch break at different time intervals; and employees whose work places are located in the first floor visit meetings only once per week to ensure the presence of the security at the building entrance.

We were also able to determine anomalies when employee visits the zone in the unusual time, when they do not return to their working place, forget using proximity switch when leaving the building. In many cases analysis of the stuff interaction would improve understanding of the anomaly origin. The BandView model can be helpful in understanding possible interaction between stuff members. But it works only when the set of employees displayed is limited and not exceeding 10–15 persons, otherwise it is very difficult to spot meetings of the co-workers.

Therefore, one of the primary tasks of the future work is the implementation of the visual analytics techniques allowing investigation of the interactions both individual and group between the stuff members in dynamics. Another direction of the future work is concerned with analysis of the data obtained from the different type of sources. For example the logs of the operating system such as login/logout events, keyboard events provide evidences that employee is at his working place. Readings from the building sensors such as building heat-ventilation system may also explain the anomalous behavior of the employee. The correlation of these data needs elaboration of the new visual analytics techniques considering the character of the source logs.

5 Conclusions

In the paper we proposed the visual analytics approach to detection of spatiotemporal patterns and anomaly in employees' movement. The key elements of the approach are interactive SOMs used to detect groups of employees and days with similar movement patterns, and heat map used to detect anomalies. They are supported by graph based and stacking based visualization technique to present spatial and spatiotemporal patterns and anomalies. We presented core interaction techniques linking all visual displays and supporting analysis process.

To illustrate our approach, we used data set provided by the VAST Challenge 2016. In the paper we discuss result obtained, and define future directions of work devoted to the enhancement of the prototype, elaboration of visualization techniques and usability evaluation of the proposed visualization analytical system.

Acknowledgement. This research has been supported by grant of the RFBR # 16-07-00625.

References

1. Millonig, A., Maierbrugger, M.: Identifying unusual pedestrian movement behavior in public transport infrastructures. In: Proceedings of Movement Pattern Analysis Workshop (MPA2010), pp. 106–110. Zurich (2010)
2. Lerman, Y., Rofo, Y., Omer, I.: Using space syntax to model pedestrian movement in urban transportation planning. *Geogr. Anal.* **46**(4), 392–410 (2014)
3. Pan, X., Han, C., Dauber, K., Law, K.: A multi-agent based framework for the simulation of human and social behaviors during emergency evacuations. *AI Soc.* **22**, 113–132 (2007)
4. Vast Challenge Homepage. <http://vacommunity.org/>. Accessed 10 Mar 2017
5. Kisilevich, S., Mansmann, F., Nanni, M., Rinzivillo, S.: Spatio-temporal clustering: a survey. In: Data Mining and Knowledge Discovery Handbook, pp. 855–874 (2010)

6. Andrienko, N., Andrienko, G.: Visual analytics of movement: an overview of methods, tools and procedures. *Inf. Vis.* **12**(1), 3–24 (2013)
7. Novikova, E., Kotenko, I.: Analytical visualization techniques for security information and event management. In: Proceedings of 21st Euromicro International Conference on Parallel, Distributed, and Network-Based Processing, pp. 519–525 (2013)
8. Schreck, T., Bernard, J., Von Landesberger, T., Kohlhammer, J.: Visual cluster analysis of trajectory data with interactive Kohonen maps. *Inf. Vis.* **8**(1), 14–29 (2009)
9. Andrienko, G., Andrienko, N.: Exploration of massive movement data: a visual analytics approach. In: Proceedings of 11th AGILE International Conference on Geographic Information Science (2008)
10. Guo, H., et al.: Tripvista: triple perspective visual trajectory analytics and its application on microscopic traffic data at a road intersection. In: Proceedings of IEEE Pacific Visualization Symposium (PacificVis), pp. 163–170 (2011)
11. Sander, N., Abel, J., Bauer, R., Schmidt, J.: Visualising migration flow data with circular plots. In: European Population Conference (2014)
12. Andrienko, G., Andrienko, N., Schumann, H., Tominski, C.: Visualization of trajectory attributes in space–time cube and trajectory wall. In: Buchroithner, M., Prechtel, N., Burghardt, D. (eds.) *Lecture Notes in Geoinformation and Cartography. Cartography from Pole to Pole*, pp. 157–163. Springer, Heidelberg (2014)
13. Guo, C., et al.: Dodeca-rings map: interactively finding patterns and events in large geo-temporal data. In: IEEE Symposium on Visual Analytics Science and Technology, pp. 353–354 (2014)
14. Kohonen, T., Honkela, T.: Kohonen network. http://www.scholarpedia.org/article/Kohonen_network. Accessed 10 Mar 2017
15. Ultsch, A.: Self-organizing neural networks for visualization and classification. *Information and Classification*, pp. 307–313 (1993)